Author:
Lazarou, Luke A J

Title:
The Population Genetics and Heritability of Measures of Immune Function in Wild House Mice, Mus musculus domesticus

The Population Genetics and Heritability of Measures of Immune Function in Wild House Mice, *Mus musculus domesticus*

By Luke Lazarou

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

School of Biological Sciences, September 2019

Word Count: 40,662

Abstract

The laboratory mouse, *Mus musculus domesticus*, is the workhorse of biological research, and have provided most of our understanding of the mammalian immune response. In contrast, the immune state and immunogenetics of wild animals have been hardly investigated at all.

To address this deficit in understanding, this study investigated the population genetics and immunogenetics of wild mouse populations. To do this, the population genetic structure of 400 wild house mice sampled from southern Britain was determined by analysis of over 1000 single nucleotide polymorphism (SNP) markers. These results showed that wild house mice exist in genetically distinct sub-populations and that the strength of this subdivision is substantially higher than has previously been reported in wild house mouse populations. The sites sampled in this study can be considered as distinct sub-populations with limited geneflow among them. Surprisingly isolation-by-distance is only in effect among the sub-populations that are located within a few kilometres of one another, around Bristol. The population structure is further supported by a nearest-neighbour joining tree analysis which exhibits strong clustering of mice within a sub-population but low similarity between sub-populations and in addition identified the presence of migrant mice. This was backed up by a STRUCTURE analysis.

The heritability of measures of immune function is debatable due to many studies reporting contradicting calculated values, with variation between species and study populations. Very few studies have examined heritability in wild mammal populations, except humans. Here the heritability of 18 measures of immune function were calculated, first for all mice sampled and secondly for mice at a single location. Like in previous studies there was variation between the two analyses performed. However, three immune phenotypes were consistently significantly heritable IgG, IgE and IL-12p70, suggesting that these genes are under strong genetic influence.

To investigate the immunogenetics of these populations, the same mice were genotyped at loci whose products have immunological function, specifically for 14 SNPs in 13 loci. The population genetics of these loci showed that there was substantial genetic diversity within them. As expected, the genes sequenced in the mouse H2 locus were highly polymorphic with a high percentage of the reported SNPs being non-synonymous. Surprisingly several immune related genes that are not part of the H2 locus were also polymorphic, specifically IL-1b. The majority of the SNPs identified through sequencing are novel in house mice. Importantly several of these polymorphisms were associated with specific immunological phenotypes, all of which have previously not been reported in house mice.

Together this work provides a novel analysis of the population genetics and immunogenetics of wild mouse populations demonstrates the need to consider that wild house mice in research. These mice harbour potentially important novel genetic variants and phenotypes which can aid in further developing house mice as model organisms for human disease and research.

Acknowledgements

The completion of this thesis has not been easy, and it would not have been possible without the support of various people. Firstly, Professor Mark Viney who has not only guided me through the PhD process and kept me on track when I wobbled. I would also like to sincerely thank NERC for funding my PhD project.

Dr Steve Paterson and Dr Andrew Turner have been of great assistance in providing some technical support and helping me make sense of the mountain of data. In addition, Dr Stephen Abolins, without whom there would be no mice or data. Dr Gibran Hemani was crucial in assisting with specific analyses and has the patience of a saint! I also need to say a very special thank you to Barney, we started this journey together and we have shared our ups and downs with each other.

Finally, I would like to thank my family and particularly, Tessa and “my David” who have had to deal with numerous frustrated phone calls and stroppy moods and provided the food to keep me going.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Contents

List of Tables

List of Figures

Glossary of Abbreviations

129X1/Sv	House mouse strain
A/J	House mouse strain
AIC	Akaike Information Criterion
AMOVA	Analysis of Molecular Variance
ANA	Anti-nuclear antibodies
B220+	B cells
BALB/cJ	House mouse strain
BalB/cJ	House mouse strain
C3H/HeJ	House mouse strain
C3H/HeJBir	House mouse strain
C57BL/10ScCr	House mouse strain
C57BL/6	House mouse strain
C57BL/6J^{5}	House mouse strain
CD11R1	Cluster Differentiation 11 Receptor 1
CD11s	Cluster Differentiation 11 Soluble
CD19+	Cluster Differentiation 19 positive
CD3/CD28	Cluster Differentiation 3 and 28
CD4+	Cluster Differentiation 4 positive
CD40lg	Cluster Differentiation 40 Ligand
CD69	Cluster Differentiation 69
CD8+	Cluster Differentiation 8 positive
CpG	5'—C—phosphate—G—3' DNA Sequence
D'	Normalised Linkage Coefficient
DBA/2	House mouse strain
DCs	Dendritic cells
DNA	Deoxyribonucleic Acid
DZ	Dizygotic
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked Immunoabsorbant Assay
EoE	Eosinophilic Esophagitis
Es	Esterase
ES-1	Esterase-1
ES-2	Esterase-2
F4/80)	EGF-like module-containing mucin-like hormone receptor-like 1
FASTA	FASTA file format

F_{IS}	Fixation index: measure of inbreeding within a sub-population
F_{IT}	Fixation index: measure of inbreeding within a total population
F_{ST}	Fixation index: measure of genetic differentiation among sub-populations
G TCA	(v1.25.3)
Gata	GATA -binding protein
GGGT	Universal - 32 1536 plex Assay kit
GWAS	Genome Wide Association Studies
Gzmb	Granzyme B
h^2	Narrow Sense Heritability
H2-Aa1	Class 2 Histocompatability 2 Aa1
H2-Ab1	Class 2 Histocompatability 2 Ab1
H2-D1	Class 1 Histocompatability 2 D1
H2-Eb1	Class 2 Histocompatability 2 Eb1
H2-K1	Class 1 Histocompatability 2 K1
H2-locus	Histocompatability 2
H2D1	H2-locus class 1 genes
H2K1	H2-locus class 1 genes
Hb	haemoglobin
Hbb locus	Haemoglobin Beta
H_S	Frequency of heterozygotes within a sub-population
H_T	Frequency of heterozygotes in a population
HWE	Hardy-Weinberg Equilibrium
IBD	Isolation-by-distance
IBDWS	Isolation by distance web service
IFN-γ	1 instance of this without hyphen
IgA	Immunoglobulin A
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgL	Immunoglobulin L
IgM	Immunoglobulin M
IL-1ra	Interleukin 1 receptor antagonist
***Il10* / IL-10**	Interleukin 10
***Il10* / IL10**	Interleukin 10
***Il12b* / IL12b**	Interleukin 12b
***Il12p40* / IL-12p40**	Interleukin 12p40
***Il12p70* / IL-12p70**	Interleukin 12p70
***Il13* / IL-13**	Interleukin 13
***Il17a* / IL17a**	Interleukin 17a
***Il1a* / IL-1a**	Interleukin 1a

***Il1b* / IL-1b**	Interleukin 1b
***Il1b* / IL1b**	Interleukin 1b
***Il2* / IL-2**	Interleukin 2
***Il4* / IL-4**	Interleukin 4
***Il5* / IL-5**	Interleukin 5
***Il6* / IL-6**	Interleukin 6
K	Cluster number (STRUCTURE)
K/BxN	Transgenic Mouse Strain
Kb	kilobases
KLH	Keyhole Limpet Hemocyanin
KLRG1	Killer cell lectin-like receptor subfamily G member 1
L_3	larval stage 3 (subscript)
LD	Linkage Disequilibrium
LPS	lipopolysaccharide
LU	London Underground
Ly49	killer cell lectin-like receptor subfamily A
Ly6G	Lymphocyte antigen 6 complex locus G6D
m.y.a.	million years ago
Mb	megabase
MHC	Major Histocompatability Complex
mRNA	Messenger Ribonucleic Acid
mtDNA	Mitochondrial Deoxyribonucleic Acid
Myo1 / MYO1A	Unconventional Myosin 1a
MZ	Monozygotic
NK	Natural Killer Cells
NKp46	Natural cytotoxicity triggering receptor 1
NNJ	Nearest-Neighbour Joining
NOD/Lt	House mouse strain
PAMP	Pathogen-Associated Molecular Pattern
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PG	Peptidoglycan
PHA	Phytohaemagglutinin
poly I:C	Polyinosinic:polycytidylic acid
RFLP	restriction fragment length polymorphisms
SJL	House mouse strain
SNP	Single Nucleotide Polymorphism
Tac2	tachykinin 2
TH1	T Helper Cells 1
TH2	T Helper Cells 2

TLR2	Toll-like Receptor 2
TLR2$_{c2}$	Toll-like Receptor 2 Cluster 2
Tlr3	Toll-like Receptor 3
TLR4	Toll-like Receptor 4
***Tlr9* / TLR9**	Toll-like Receptor 9
TNF-α	Tumor Necrosis Factor Alpha
UTR	Untranslated Region
WBC	White Blood Cell
$\gamma\delta$	T-cell Sub-set

Chapter 1: Introduction

General Introduction

Research using the house mouse has been at the forefront of advances in biomedical research for almost a century and much credit is owed to this species for giving us the immunological and genetic understanding of mammals we have today (Viney *et al.*, 2015). The laboratory study of house mice has been aided by using reductionist methods to reduce confounding effects in experiments, and these include the use of defined genetic stocks (inbred strains), sterile conditions and tightly controlled environmental factors (Viney *et al.*, 2015). House mice lend themselves to being an excellent model species due to their shared mammalian biology to humans, their small size, and rapid breeding rate.

Immune function, the ability of a host's immune system to recognise and neutralise an invading pathogen can vary among, and within species. Variation also exists in the genes associated with immune function (immunogenetics) in many species and wild populations (Maizels and Nussey, 2013). In comparison to wild animals, domesticated and laboratory bred strains of animals show little within-strain heterogeneity in immunological traits and are the primary research tool utilised in immunological research (Viney and Riley, 2014). However recent studies of wild house mice have highlighted the significant differences in measures of immune function that exist between these wild mice and their laboratory bred counterparts (Thia and Trapani, 2007; Abolins *et al.*, 2011; Boysen *et al.*, 2011). Further, studies of wild-derived inbred house mice (the offspring of wild caught mice bred with standard laboratory strains) have suggested the presence of a yet unidentified molecule that is able to influence pro-inflammatory signalling pathways, and has not previously identified among standard laboratory mouse strains (Stephan *et al.*, 2007). The emerging fields of eco-immunology and immunogenetics have highlighted the differences that exist between our well-studied domesticated and laboratory bred research models and their wild counterparts. As house mice are the primary mammalian genetic and immunological research model used today, the study of wild populations of this species can

considerably improve the validity of data obtained from studies of laboratory mice. Many studies of “wild animals” often involve elements of artificial conditions and do not truly represent the wild populations.

Eco-Immunology

Our immune system is essential in preventing pathogenic infections. These infections can dramatically affect the health of a host, reducing its survival chance and importantly its ability to reproduce. It therefore makes sense that individuals best adapted to resist infections have greater reproductive success and pass on these alleles to the next generation. Genes play a critical role in determining the response of the immune system as many proteins are involved, such as cytokines, chemokines, cell surface markers and receptors and many more. Pathogens are a selective pressure on these genes and significantly affect which alleles are present in a population. As pathogens vary greatly, genetic variation in a population increases the chance that some individuals in a population are resistant of infection. The pathogen-immune function relationship is complicated by the fact that hosts can be infected by multiple pathogens and different pathogens require different responses that can often conflict with each other. Additionally, the maintenance of efficient immune function is costly. Mounting an immune response requires considerable energetic and nutritional investment and this is often at the detriment of other physiological processes such as growth and reproduction (Festa-Bianchet, 1989; Lazzaro and Little, 2009). Therefore, a trade-off occurs with many competing environmental selective pressures further adding to genetic variation in a wild population. Understanding the genetic variation of a population gives us an insight into the evolutionary forces acting on a population. More specifically variation in genes with immunological functions can inform us about the pathogenic pressures that populations face.

A major shortcoming of laboratory mouse studies is the lack of natural context that mice face when immunologically challenged; for example, animals are often only challenged with a single exposure to an infection. However, the development of genetic and immunological resources in laboratory mice allows for the application of these resources to wild populations

Eco-immunology is the study of the immune system of a host organism in the context of its natural environment. In contrast to laboratory-based immunology which reduces all environmental variation, eco-immunology looks at these interactions and the underlying mechanisms that influence immune function.

In laboratory house mouse strains, the immune responses to immunological challenges are highly ubiquitous (Lesher *et al.*, 2006) as are the genotypes of these individuals and the sterile environments they live in. The benefits of laboratory studies are clear; they reduce experimental noise and variance leading to stronger statistical effects and a clearer aetiology of aspects of immune function. However the immune system of vertebrates evolved in variable environments where infections were frequent and therefore selective pressures on immune function were greater in comparison to laboratory house mice (Viney and Riley, 2014). Therefore, our continued study of laboratory immunology, particularly of the house mouse, may not represent a host's immune function in a realistic context (Viney and Riley, 2014). In contrast to laboratory bred animals significant heterogeneity exists among wild hosts in their ability to resist infections, with the minority of a population often harbouring the most intensive infections (Hayward, 2013). By studying laboratory bred animals in sterile environments we risk overlooking important novel variants and interactions that cannot be observed in a traditional experimental setting (Stephan *et al.*, 2007; Tschirren *et al.*, 2013). Eco-immunology seeks to take our knowledge of immunology from laboratory studies and apply this to natural populations in an attempt to overcome the constraints of studying laboratory bred animals and improving the ecological validity of immunological research (Turner, 2010).

The environments wild animals live in can affect their immune function in a number of ways. Wild animals are likely to have experienced immunological challenges throughout their life. These past infections can have significant effects on a host's immune system, potentially polarising the immune responses to immunological challenge (Freeman, Jr., 2009; Cassetta *et al.*, 2011). Animals can also harbour co-infections at the point of experimental challenge that may influence their response to

immunological stimulation (Lello *et al.*, 2004; Pedersen and Fenton, 2007; Telfer *et al.*, 2010; Dietze *et al.*, 2016). Also, laboratory bred animals are often maintained with unlimited nutritional resources, whereas wild animals may experience trade-offs in immune function due to limitations of food and other physiological processes such as reproduction (Walford *et al.*, 1973; French *et al.*, 2007; Graham *et al.*, 2010). However intrinsic factors such as genotype may also contribute significantly to the observed heterogeneity in measures of immune function in wild animals.

Laboratory house mice have been selectively bred for generations resulting in almost complete genetic homozygosity and therefore they have very low within-strain immunological variation. However the ability of lab mice to be bred and selected for high and low immune responses to immunological stimulation suggests a strong genetic component to immune function (Sitepu and Dobson, 1982). More specifically, immunological comparison of strains of laboratory mice, cell counts and antibody titres, as well as differences in infection susceptibility, show phenotypic variability (reviewed by Sellers *et al.*, 2012), which is further evidence that inter-individual differences in genotype contributes to variation in immune function. Quantitative genetic studies of humans suggest an average heritability of immunological traits is 41% (Orrù *et al.*, 2013). While laboratory bred animals may not exhibit much, if any, within strain genetic variance, the genetic composition of wild mice is comparatively substantially different (Thia and Trapani, 2007). Laboratory house mice have been derived from a limited gene pool and therefore genetic variation among wild mice may be an important source of novel immunological variants and mechanisms, bridging the gap between wild and laboratory immunological understanding.

Specific immunological differences between wild and laboratory-bred house mice have been highlighted in several recent studies (Stephan *et al.*, 2007; Thia and Trapani, 2007; Abolins *et al.*, 2011; Boysen *et al.*, 2011). Importantly, these studies have also shown that a much greater amount of inter-individual variation exists among wild house mice compared to their laboratory counterparts (Thia and Trapani, 2007; Abolins *et al.*, 2011; Boysen *et al.*, 2011). Abolins *et al.* (2011) examined various cellular and humoral measures of immune function among wild house mice in south

west England (Abolins *et al.*, 2011). The results showed that wild mice had significantly higher anti-KLH (keyhole limpet hemocyanin) IgG and IgE antibody titres, and greater avidity of anti-KLH IgG following intra-peritoneal immunisation with KLH antigen. Following stimulation with concanavalin A (Con A) proportional, antigen-nonspecific, cell counts of splenic leukocytes showed significant changes, particularly in the increase for CD4+ T helper cells and decrease in CD8+ Cytotoxic T cells and B220+ B cells in wild mice (Abolins *et al.*, 2011). Additionally, there was greater variation among the wild mice, than among the laboratory bred mice, in anti-KLH IgG and IgE blood titres and in CD4+, dendritic and macrophage cell counts (Abolins *et al.*, 2011). Similarly, although functional effects were not examined, comparisons of the amino acid sequences of granzyme b in wild and laboratory house mice showed that a much higher amount of sequence variation exists among wild and wild-derived house mice compared to their laboratory counterparts (Thia and Trapani, 2007). Among the 13 laboratory bred strains of mice studied only 1 strain showed any variation (DBA/2) having 2 amino acid differences compared to the other strains (Thia and Trapani, 2007). In comparison, a very large amount of variation was observed among wild and wild-derived house mice and the specific allozyme of granzyme B from the inbred mice was identified in only one 'clan' of wild mice (Thia and Trapani, 2007). This suggests that the laboratory bred mice used in immunological studies represent a very limited sample of the genetic and immunological variation that exists in the species as a whole. Large amounts of phenotypic variation have also been observed in innate immune responses of wild-derived inbred house mice (wild caught mice bred with a laboratory strains and outcrossed), to stimulation with polyinosinic-polycytidylic acid (poly(I:C)) (Stephan *et al.*, 2007). Wild mice showed considerable variation in the amount of TNF-α produced in response to poly(I:C) stimulation which was also associated with a mutation in the *Tlr3* gene (Stephan *et al.*, 2007). Additionally, this study suggests the presence of, a yet unidentified, immunomodulatory molecule that is able to alter proinflammatory pathways, again highlighting the importance of studying immunology in wild populations where there may be unique variants important for mechanistic understandings of immune function. However, perhaps the most significant difference that has been recorded to date is seen in the function of natural killer (NK) cells in wild and laboratory mice. NK cells are cytotoxic lymphocytes

that are part of the innate immune system. In comparison to human NK cells, cultured mouse NK cells respond less strongly to antigenic stimulation (Bergman *et al.*, 2000). However, wild house mice have shown a significantly higher number of NK cells ($NKp46^+$ / $CD3^+$) in peripheral lymph nodes (Tschirren *et al.*, 2013). Further, the NK cells of these mice also exhibited a unique phenotype not previously documented in laboratory studies. NK cells had increased expression of Gzmb, CD69 and KLRG1 in splenic populations (Boysen *et al.*, 2011). This study clearly indicates that pre-exposure or priming of natural killer cells is essential to produce adequate effector cells in response to antigenic challenge, and this further demonstrates the importance of context in determining immunological phenotype. However, the exact mechanisms that underlies this novel immune phenotype is not get known and genetic variants may play an important role, but this cannot currently be investigated using the existing inbred stocks of house mice due to the lack of relevant phenotypic variation in these mice and their genetic homogeneity.

A major shortcoming of laboratory mouse studies is the lack of natural context that mice face when immunologically challenged, often challenged with a single exposure to an infection. However, the development of genetic and immunological resources in laboratory mice allows for the application of these resources to wild populations of house mice which may prove an abundant resource for novel interactions in a natural context, bolstering the validity of existing mechanistic understandings.

Immunogenetics

Immunogenetics is an expanding field that aims to link variation in immune function to genetic variants in order to better understand the relationship between genotype and immunological phenotype. As discussed above, there is a large amount of evidence that suggests a host genotype plays a significant role in determining its immune function, and genes associated with the immune system are among the most variable across species (Hughes and Hughes, 1995; Meyer and Thomson, 2001). This variation is likely a result of the selective pressures that are exerted on the immune system by the environment (Viney *et al.*, 2005). Traditionally, immunogenetics has been studied in the context of a laboratory, and compared strains of inbred animals, linking genetic variances among strains to their phenotypic differences. However, the limited gene pool of these animals limits the capacity to identify the effects of novel genetic variants and the mechanisms involved. Using an eco-immunogenetic approach may identify novel phenotype-genotype associations and novel alleles not present in laboratory mouse populations.

Along with inbred laboratory strains of animals, domesticated livestock have also experienced intense artificial selection in search of desirable phenotypes such as faster growth and greater body weight. Infections still pose a significant threat to livestock and understanding the aetiology of immune function among wild animals may assist in guiding practices in animal husbandry and selection. The Soay sheep of the St. Kilda archipelago in Scotland are a free-living, unmanaged population that have been intensively studied since the 1950s. Several immunogenetic studies have been conducted on this population (Paterson *et al.*, 1998; Coltman *et al.*, 2001; Brown *et al.*, 2013). Microsatellite variants within the interferon gamma gene are significantly associated with the faecal egg counts of yearling sheep (Coltman *et al.*, 2001). In addition, both the survival of juvenile sheep against nematode infection and the resistance of sheep to these infections is associated with differences in the major histocompatibility complex (MHC) (Paterson *et al.*, 1998), a locus that has been at the forefront of immunogenetic investigation for many years (Acevedo-Whitehouse and Cunningham, 2006). More recently, a candidate gene approach has been utilised on this population to examine associations between a set of preselected genes-of-

interest and measures of immune function (Brown *et al.*, 2013). This study did not find any significant association between any single candidate gene and any of the immunological measures. However, in this study several SNPs explained variation in multiple immune traits even though very few direct immunological measures were taken from these sheep (Brown *et al.*, 2013). This suggests that a candidate gene approach may still be applicable to immunogenetic studies in wild populations but the selection of candidate genes and the parameters used for immune function need to be carefully considered.

Like mice, voles are similarly a small, abundant rodent species that are less commensal than house mice. Additionally, the orthologous nature of the immune system among these species allows for genetic and immunological resources developed in house mice to be applied to wild populations of voles (Turner *et al.*, 2011; Tschirren *et al.*, 2013). Immunogenetic studies of wild voles have recently found genetic associations to resistance to infections and to immunological parameters of immune function. In bank voles (*Myodes glareolus*) variants of the TLR2 gene (part of the $TLR2_{c2}$ cluster) was found to be significantly associated with resistance to infection with *Borrela afzelii* (Tschirren *et al.*, 2013). In total 15 different haplotypes were identified among the sampled voles, demonstrating the high level of variation that exists in a wild population for a key antigen-sensing receptor. Genetic associations have also been identified in field voles (*Microtus agrestis*) to infection resistance and mRNA expression of cytokines (Turner *et al.*, 2011). Specifically, polymorphisms in *Il1b*, *Il2* and *I12b* cytokine genes were significantly associated with the mRNA expression of IL1b, Gata96h (*Il1b* gene), IL10 (*Il2* gene), IL1b and IL2 (*Il12b* gene) (Turner *et al.*, 2011). While these studies show significant immunological genetic associations, they are also limited by the measures of immune function that they examine. Neither study directly measure immune function at an effector level of protein expression or cell counts, and despite Turner *et al.* (2011) examining the mRNA expression of immunological measures this may not directly translate to the protein expression level. A hindrance of many immunogenetic studies of wild populations is the availability of genetic and immunological resources. While voles are similar enough to house mice that some resources can be applied from the laboratory, the differences between these species

may inhibit accurate measuring of parameters of immune function. Further, for most non-model species no immunological resources or genetic database exist making immunogenetic research of wild populations challenging.

Historically, immunogenetic studies of mice have been indirect, often inferring genetic variation from the use of different inbred strains and associating these to immunological differences, predominantly resistance to infectious organisms (Pepose and Whittum-Hudson, 1987; Williams and Timoney, 1994; Medina *et al.*, 2001). More recent studies have shifted focus to the mechanistic underpinning of resistance to infection by directly examining immunological measures of immune function such as cell counts and antibody titres (Mostafavi *et al.*, 2014). However, this is still a developing field, with few studies completed and the gap between laboratory immunogenetics and wild populations is vast.

The challenges that immunogenetic studies in wild populations face are vast due to the complexity of the relationship between organisms and their environment, not to mention the limited tools that are available for non-model species. The variation that exists in wild populations makes them of research interest for identifying novel phenotypes, alleles and interactions not seen in laboratory populations. However, the ability to identify causative loci is hindered by the confounding intrinsic and extrinsic variable that affect these organisms. Rather than applying a reductionist laboratory-based methodology to these wild populations, studies of wild organisms need to accommodate for the variation that exists. Therefore, additional data is required in comparison to laboratory-based genetic association studies, as statistical analyses must take into account the effects of these environmental factors. Immunogenetic studies are also limited by the tools available to study both the genetic and immunological components. Very few species have had their genomes sequenced in detail, which makes it difficult to select loci for investigation. Additionally, the immunological reagents required for such studies are often species-specific, restricting the ability of studies to directly measure immunological measures of immune function. However, the house mouse is one species that addresses these issues. The house mouse has a well characterised genome thanks to extensive genetic

studies in the laboratory and immunological tools such as ELISA kits and flow cytometry, optimised for laboratory mice, can be directly applied to their wild counterparts.

Very few studies have attempted to link immune function and genetics in a wild context. Thia and Trapani (2007) successfully identified allelic differences in the immune related gene Gzmb gene in wild house mice (Thia and Trapani, 2007). However, this study did not examine the functional immunological consequence of these genetic differences, nor if it affected immunological parameters of immune function. In wild-derived house mice significant genetic associations between variants of the TLR3 gene and hyposensitivity to antigenic stimulation with poly (I:C) have also been identified (Stephan *et al.*, 2007), but again there were no immunological measurements made, meaning that we do not know if an immunological mechanism was responsible for this hyposensitivity. Despite no direct immunogenetic studies having been conducted on wild house mouse populations these studies suggest that the extensive immunological variation among wild house mice (Lochmiller *et al.*, 1993; Abolins *et al.*, 2011; Boysen *et al.*, 2011) may be explained, at least in part, by the substantial amount of genetic variation in wild populations.

The Mouse Model

Evolution and Dispersal

In many cases it is not possible, or simply unethical, to study aspects of the immune system in humans directly. Humans cannot be deliberately infected with many pathogenic infections and require treatment after infection, preventing the long-term study of the infection pathology. Additionally, invasive measures of immune function, such as lymphocyte extraction from spleens ca not be performed on living humans, therefore animal models are required to study immune function in mammals. The house mouse is an attractive model due to its small size, rapid life cycle (for a mammal) and orthologous immune system to humans (Phifer-Rixey and Nachman, 2015). The house mouse is the predominant mammalian model used in immunological research today (Viney and Riley, 2014). Originating in the north Indian sub-continent, based on mitochondrial DNA (mtDNA) sequences and restriction fragment length polymorphisms (Boursot *et al.*, 1996; Din *et al.*, 1996), the house mouse diverged from *Mus spretus* ~ 1.7 million years ago (m.y.a) (Suzuki *et al.*, 2013). The commensal nature of the house mouse meant that the species was able to quickly capitalise on the migration of their human partners. Mitochondrial DNA (mtDNA) haplotypes closely mirror the pattern of human colonisation (Jones *et al.*, 2011; Gabriel *et al.*, 2015; Hardouin *et al.*, 2015). For example, Cyprus, an island with a long history of human colonisation by different civilisations throughout history has a highly diverse set of mtDNA haplotypes present among its wild house mice, compared to Gran Canaria, colonised relatively recently, where only two mtDNA haplotypes are present (Hardouin *et al.*, 2015). Today, numerous subspecies of *Mus musculus* have been identified with the most well characterised, *Mus musculus musculus*, *Mus musculus domesticus* and *Mus musculus castaneus* estimated to have diverged from one another between 0.37 – 0.4 m.y.a (Suzuki *et al.*, 2013). *M. musculus domesticus* is the subspecies of house mouse found in western and northern-Europe. The expansion and dispersal of *M. domesticus* into Europe is estimated to have occurred roughly 12,000 years ago during the human Iron age (Bonhomme *et al.*, 2012). In the Britain Isles, mtDNA sequences of wild house mice indicate that mice living in Ireland and Scotland have a shared ancestry with Norwegian mice suggesting they were

introduced by the Viking colonisation of these regions (Jones *et al.*, 2011). Mice from the rest of the British Isles share a clade (i.e. are a genetically similar grouping) with mice from northern France. The mainland British-French and Viking clades are limited to north-western Europe which may reflect maritime colonisation by these mice (Jones *et al.*, 2011).

In Research

The house mouse is reported to have first been used in biomedical research in 1664 by Robert Hooke and his investigations of air pressure on organisms (Berry, 1967). In fact it has even been reported that Mendel's first experiments were conducted with house mice examining variants in their coat colour before his more famous experiments with peas (Paigen, 2003a, 2003b). However, it was not until the early 20th century that the modern use of mice in biomedical research took off. Since then house mice have played an integral role in the development and understanding of the fields of genetics, immunology and oncology (Guénet and Bonhomme, 2003).

Studies of house mice have benefited from the use of highly standardised strains of mice created from generations of inbreeding to remove genetic variation that may create experimental noise. The first inbred mouse strain (DBA/2) was created in 1909 by C.C Little based on breeding for coat colour (Russell, 1978). Many of the laboratory bred strains of house mice available today derive from a small starting population of mice bred by mouse fanciers for specific coat traits (Guénet and Bonhomme, 2003). Since then an enormous range of inbred strains have been created, each bred for specific phenotypes. While the homogeneity within a strain has allowed for a high degree of experimental reliability the restricted gene pool inhibits the validity of mouse data being applied to other species.

Genetics

The house mouse was the first non-human mammal to have its genome sequenced in 2002 (Waterson and Consortium, 2002). The development of this essential resource allowed for a rapid expansion in the use of house mice in biomedical research. Gene mapping, genetic association studies, and the creation of transgenic organisms has all been made possible, allowing for both forward and reverse genetic studies to be conducted to identify the causative genes of specific traits. Genome Wide Association Studies (GWAS) are a powerful tool in identifying genetic associations in a specific organism to a phenotypic trait, or traits, of interest. GWAS have been utilised successfully in humans and livestock to identify causative loci of disease (Gregersen, Olsson, 2010; Thompson-Crispi *et al.*, 2014; Kochi, 2016; Psifidi *et al.*, 2016) since, as in the house mouse, there is a high coverage of markers in the genomes of these species. However, the application of GWAS is limited largely to model organisms, livestock and humans where the genome is highly characterised, a large number of samples can be obtained, there is sufficient coverage of genetic markers across the genome, and phenotypes can be easily measured (reviewed Korte and Farlow 2013). Therefore, the application of GWAS to wild populations, which often do not meet these criteria, is very limited. Even in the house mouse, which does have a well characterised genome, where phenotypic traits are relatively easy to measure, and the population sizes are high, the possibility of intrinsic relatedness among individuals within a population and the confounding effect of genetic background in geographically distinct populations limits the effectiveness of GWAS (Fournier-Level *et al.*, 2011; Vilhjálmsson and Nordborg, 2013).

Commensal house mice live in close proximity to humans and therefore their biology is directly influenced by human activities. The patchy population genetic structure of house mice is likely a result of anthropogenic influences. Pest control programs mean that populations are often unstable, surviving for only a few generations. In addition, the man-made environments they inhabit often mean their exposure to pathogens varies. For example, mice living at stations on the London Underground often scavenge for food on platforms but sleep, urinate and defecate on the tracks. This behavioural shift in the mice means that the faecal-oral transmission route of

nematode infections is disrupted and as a result very few worm infections are observed in these mice (Abolins *et al.*, 2017). These interactions may suggest that the immune function of wild house mice would not truly reflect that of a wild species (Jackson, 2015). However, it is these precise attributes that make wild house mice the perfect wild population to model human immune function. Only wild house mice have experienced the similar changes in environmental and pathogenic conditions that humans have, while still remaining genetically diverse, compared to domesticated species such as dogs and cattle. In developed countries, asthma and allergies, diseases associated with the dysregulation of the immune system, are becoming more prevalent (Hadley, 2006; Lerner *et al.*, 2016) and wild house mice may provide a unique bridge to understand how our human-made environments influence a species' immune function. Further, the genetic diversity present in these wild house mice allows for not just an immunological understanding, but also a potential immunogenetic link between genotype and immune function phenotype.

Population Genetics

The field of population genetics seeks to understand the pattern and quantity of genetic variation that exists within and among populations and to determine the processes that drive this. Studies in this field initially utilised allozymes, structural variants of a particular enzyme (Petras 1967a; Selander *et al,.* 1970), and *t*-alleles, allele versions of the *T-locus*, as genetic markers (Lewontin & Dunn 1960; Anderson 1964; Petras 1967b), which later developed into the use of microsatellite loci (Dallas *et al,.* 1995; Panithanarak *et al,.* 2004), and eventually the inclusion of mitochondrial DNA sequences and bi-allelic Single Nucleotide Polymorphisms (SNPs) (Searle *et al.* 2009). Early population genetic studies typically investigated a handful of loci at a time (Anderson 1964; Petras 1967a). Using a small number of loci can significantly bias the results of population genetic studies, as the results are representative for only a small number of loci, which may be under the influence of specific selective forces. More recently there has been a shift towards an increased use of multi-locus SNP genotypic data. This transition is due to technological advances in sequencing, candidate SNP identification, and the availability of rapid genotyping platforms.

The degree of polymorphism for different loci can also affect the choice of the number of loci used in studies. For example, microsatellite loci are highly polymorphic and the use of 10-20 microsatellite loci can achieve the same discriminatory power of about 100 SNP loci (Kalinowski, 2011). However, the advantage of using SNP locus genotyping is that they are very common and evenly spaced across the genome, unlike microsatellite loci. This greater abundance of SNPs allows for the rapid genotyping of large numbers of loci at a time. Most studies using SNP data far exceed the level of discrimination typically achieved through the use of microsatellite or allozymes data.

Genetic Analyses

With the emergence of new data formats and an increase in the volume of data to be analysed new statistical methods have been developed to complement pre-existing methods. Here I will discuss some of the analytical methods commonly used with multi-locus SNP data sets.

Testing for Neutrality and Linkage Disequilibrium

Neutrality

Neutral loci are those that are not under selection, and therefore their allele frequencies are completely dependent on population dynamics. Non-neutral loci contribute to traits that are under selection such that their allele frequencies result from selective forces acting on them, or on other linked loci. In population genetic analyses it is important to use only neutral loci in analyses to prevent biasing the results. Non-neutral loci should be utilised in studies examining the effect of selective processes on the allele frequencies of loci directly associated with a specific function, such as immunity. The allele frequencies of the major histocompatibility complex (MHC) are often used to study the effects of selective processes on the genetics of immunity in mammals (Yao *et al,.* 2014). Neutral loci provide a more general insight into animals' population structure, the level of genetic variation and non-selective processes acting on populations, such as migration and genetic drift.

Linkage Disequilibrium

Loci that are in linkage disequilibrium (LD) do not randomly segregate during meiosis. This means that linked loci are inherited together more often than would be expected by chance. LD can be used in population genetic studies to determine the amount of random mating that occurs through studying the physical distances among linked loci. Weak LD among physically distant markers indicates that lots of outbreeding has occurred, whereas strong LD would suggest that the population is relatively inbred, sharing a large amount of genetic material. In inbred populations, such as inbred strains of house mice, LD can be observed among loci over one megabase (Mb) apart,

compared to outbred wild mice where strong LD is only observed with loci of less than 100 kilobases (Kb) apart (Laurie *et al,.* 2007).

Heterozygosity

Hardy-Weinberg Equilibrium

The Hardy-Weinberg Equilibrium (HWE) predicts the stable, multi-generation allele frequencies in an ideal population based on the equation:

$$p^2 + 2pq + q^2 = 1$$

The HWE exists in ideal populations that are infinitely large, that are randomly mating, and have no genetic drift, mutation or migration. In reality these conditions rarely exist in nature and in these cases the frequency of heterozygote genotypes may significantly deviate from the HWE expectation. Importantly, such deviations can be used to understand the evolutionary forces acting on a population. For example, an excess of heterozygote genotypes could result from the recent acquisition of new mutations in a population, or from positive selection for heterozygote genotypes. A heterozygote deficiency could be an indicator of non-random mating, inbreeding, or genetic drift having occurred, resulting in population subdivision.

If a large number of loci in a population significantly deviate from HWE expectations, this might be evidence for population sub-division, a phenomenon known as the Wahlund effect (Wahlund, 1928). In this situation a population that is actually subdivided, the allele frequencies of each of the two sub-populations are different, but when the population is considered as a whole it would appear to have a deficit of heterozygotes.

Wright's F_{st}

Wright's F_{ST} *(Wright, 1922)* is an inbreeding coefficient analysis that compares the observed frequency of heterozygotes within a sub-population to the expected frequency of heterozygotes in the total population. using the following equation:

$$\mathbf{F_{ST} = H_T - H_S/H_T}$$

Where H_T is the frequency of heterozygotes in the total population and H_S is the frequency of heterozygotes among individuals within a specific sub-population. F_{IS} and F_{IT} are other inbreeding coefficients, often calculated in population genetic studies. F_{IS} describes the mount of genetic variation that exists among individuals within sub-populations, whereas F_{IT} describes the genetic variation that exists among individuals across the total population.

F_{ST} values range between 0, where no genetic sub-division is present, and 1 where the population is completely sub-divided with no gene flow occurring. F_{ST} values are often presented as the proportion of genetic variance attributed to differences among sub-populations. The average total F_{ST} value for a number of loci represents the overall genetic differentiation among sub-populations. For a finer resolution of genetic sub-division among sub-populations, the between sub-population pairwise F_{ST} values can be calculated to understand the relative distance between multiple, different sub-populations.

Analysis of Molecular Variance (AMOVA)

AMOVA can partition the genetic variance in a population among sub-populations and the total population in a manner analogous to F_{ST} analyses. In addition to pairwise F_{ST} analyses, AMOVA can also be done in a hierarchical manner, to determine the proportion of genetic variance attributed to individuals within a sub-population, between sub-populations, and also among groups of sub-populations.

Population Structure

F_{ST} and AMOVA analyses can be used to indicate whether a population is genetically sub-divided or otherwise structured, but these do not give a very fine resolution of the population structure. Further to this, to analyse large scale genotypic data sets it is necessary to use computational algorithms to further explore population structure. These finer scale analyses can be done using other approaches including that employed by the programme STRUCTURE and through the construction of a nearest-neighbour joining (NNJ) tree.

STRUCTURE

STRUCTURE (Pritchard *et al.*, 2000) is a Bayesian-based algorithm model for determining the true number of sub-populations that exist in a sample and to visualise the population structure. While F_{ST} is able to test the strength of genetic differentiation between two pre-defined sub-populations, STRUCTURE is able to identify and define sub-populations among individuals. Under the admixture model STRUCTURE (Pritchard *et al.*, 2000) uses a Bayesian type algorithm to assign portions of an individual's genotype to pre-defined genetic clusters. The number of clusters (K) is controlled by the user and then pre-defined by the programme from the genotypic data allowing hypotheses of the number of clusters to be tested. In these analyses portions of an individual's genotype may be derived from multiple clusters, which indicates mixed ancestry and gene flow between sub-populations. If individuals from different sub-populations are found to mainly derive their genotypes from separate clusters, this suggests that little gene flow occurs between these sub-populations.

The results of these STRUCTURE analyses can be graphically visualised using CLUMMP (Jakobsson and Rosenberg, 2007) and DISTRUCT (Rosenberg, 2004) or the program CLUMPAK (Kopelman *et al.,* 2015), which compile multiple analysis runs and which plots the data in a consensus graph. Examination of these plots through increasing values of K can be informative of the hierarchical population structure revealing the order in which sites differentiate from one another. The value of K that best explains the data, and therefore that suggests the likely number of sub-populations, can be

calculated using the posterior probabilities of each run calculated using the Evanno method (Evanno *et al.,* 2005).

Nearest-Neighbour Joining Tree

Phylogenetic trees can be used to determine the evolution and ancestry of species and individuals based on similarities in their genomes, generally DNA sequences. However, the data generated from multi-locus SNP genotyping are not part of a DNA sequence and do not necessarily share an evolutionary or selective ancestry with one another, therefore phylogenetic reconstruction through maximum parsimony or maximum-likelihood trees are not appropriate. Instead, a nearest-neighbour joining tree can be applied to these data in order to interpret and display the distribution of genetic variation among individuals of a population. This is done by the plotting of a tree based on the pairwise genetic distance among individuals which is calculated as the number of nucleotide differences among individuals. From this tree patterns in the branching and positioning of individuals can be used to indicate the relative similarity of individuals and groups of individuals, and to identify anomalous individuals, such as migrants, for further investigation.

Isolation-by-distance

Individuals in a population may become reproductively isolated due to the geographical distances among them (Wright, 1946), resulting in population sub-divisions. To test for this an Isolation-by-distance (IBD) analyses can be used. IBD analyses look for correlations between genetic and geographical distances. It can be assumed that as geographical distance between individuals increases, then the likeliness of them mating (and so sharing alleles) decreases. A Mantel test of correlations between genetic and geographical distance can then be used to determine whether individuals are genetically isolated by distance. These analyses can be conducted using programs such as Isolation-by-Distance Web Service (IBDWS) (Jensen *et al.*, 2005).

Population Genetics of Domestic and Wild Animals

Domesticated species are of significant economic interest to humans. Through generations of selective breeding the majority of domesticated animals exist as characterised breeds that have been selectively bred for specific traits. These breeds are maintained through further selective breeding meaning that gene flow among breeds is virtually non-existent. Recent population genetic studies have assessed the amount of genetic sub-division that exists among domesticated animal breeds.

Pigs were originally domesticated from wild boars, and then selectively bred for traits such as growth and litter size, to increase meat yield, and this process produced the commercial pig breeds available today. Several studies have examined the genetic differentiation that is present among breeds of domesticated pigs and also between domestic pigs and wild boars. Comparisons of Chinese indigenous and commercially available European breeds of pigs have found that a moderate amount of genetic differentiation among them exists, with an F_{ST} value of 0.18 among the various breeds, and that the indigenous pig breeds and European pig breeds are distinct lineages (Fan *et al.,* 2002). This amount of genetic differentiation is similar to that observed between island and mainland populations of wild boars in both Japan and Italy (Murakami *et al.,* 2014; Iacolina *et al.,* 2015), with F_{ST} values of 0.178 and 0.126 - 0.138, respectively. Both the Japanese and Italian studies also compared wild boars to domesticated pig breeds, with F_{ST} values found to be much higher in the Japanese study, 0.471 and 0.560 (Murakami *et al.,* 2014), compared to the Italian study, 0.169 (Iacolina *et al.,* 2015). These results suggest that the wild Japanese pigs have been genetically isolated for longer from the domesticated pig breeds they were compared against, than the wild boars and domesticated pigs in the Italian study. However, the parity among these studies is limited because the domestic pig breeds used in the studies are not the same and therefore the differences between them are not directly comparable.

Cattle are also of high economic interest to humans. A population genetic study comparing 80,000 SNP loci in two populations of zebu cattle (native to South-east Asia) and semi-domesticated gayal cattle found high F_{ST} values of 0.33 (Uzzaman *et al.,* 2014). This indicates very strong genetic sub-division between these two breeds of cattle, likely due to the selective breeding practices of farmers. Zebu cattle, which are less domesticated than gayal cattle, had higher frequencies of heterozygotes compared to the gayal cattle, suggesting that these cattle are indeed more outbred (Uzzaman *et al.,* 2014).

Chickens are one of the most widely distributed and consumed livestock species. Chickens have been bred for several distinct traits, such as increased body mass and rapid growth for meat production, increased egg laying for food production, and for their plumage for use as pets. A large-scale study of chicken populations in Africa, Europe and Asia used 29 microsatellite markers to analyse the genetic differentiation among 113 indigenous chicken breeds, 3 red jungle fowl populations and 9 commercial chicken lines (Lyimo *et al.*, 2014). This study found that in the African and Asian breeds, and in the red jungle fowl populations, the observed level of heterozygosity was high, meaning that these breeds and populations are relatively outbred compared with their European counterparts. The highest F_{ST} values were observed among comparisons of European breeds and commercial chicken lines in contrast with African and Asian chicken breeds. Further analysis using STRUCTURE found that the greatest population division was between Asian and North-west European chicken breeds. The Asian chicken breeds were relatively outbred, and therefore these results suggest that this genetic differentiation has come about through the isolation and selective breeding of the North-West European breeds with respect to the Asian breeds (Lyimo *et al.,* 2014).

Together, these analyses of different breeds of domesticated animals, both compared among breeds within a species and between domesticated breeds and wild relatives, all show consistent, notable genetic differentiation among breeds. These results are consistent with each breed having a breeding structure which is largely isolated from other breeds, and that these breeds are also genetically isolated from the relevant

wild, non-domesticated species. The extent of genetic sub-division among domesticated animal breeds can be used as a standard against which to compare wild, non-domesticated populations.

Wild Animal Population Genetics

Other studies have also analysed the extent of genetic differentiation among wild mammalian populations. A study of three putative sub-populations of dromedary camels (*Camelus dromedarius*) in Tunisia, where morphological differences suggested the existence of different sub-populations, were investigated using four microsatellite loci with a total of 26 alleles to determine whether they were genetically differentiated. These results showed that the mean F_{ST} value was 0.052, which indicates a low level of genetic differentiation. Therefore the study concluded that the putative sub-populations were only partly genetically isolated which may suggest that they starting to differentiate (Nouairia *et al.*, 2015).

Studies of eight fragmented colonies of black-tailed prairie dogs (*Cynomys ludovicianus*) calculated an F_{ST} of 0.115 among individuals. The between-colony differentiation was higher among those colonies that had been managed. In these colonies there was no isolation-by-distance, suggesting that insufficient migration among colonies occurred to counter any genetic sub-division (Daley, 1992).

The genetic sub-division of wolverines (*Gulo gulo*) sampled from five distinct sites in Canada was determined using allozymes and mtDNA haplotypes. This study found that there was a very low F_{ST} value among sites, calculated from allozymes data. However, the mtDNA haplotype comparisons gave an F_{ST} value of 0.536. This finding suggests that the gene flow among the sample sites of this study is largely dependent on males. Male migration, and therefore male dependent, gene flow limits the divergence of nuclear allele frequencies, but a low amount of female migration results in the divergence of the mitochondrial genomes (Wilson *et al.*, 2000).

A study of the Texas deer mouse (*Peromyscus attwateri)* from 12 sites in Arkansas used 15 polymorphic loci to determine the population structure of the mice. The total F_{ST} value, calculated across all mice and loci, was 0.137, which suggests a moderate amount of genetic differentiation among the sample sites. Physio-geographic regions accounted for only 1% of the overall genetic variance observed in these mice (Sugg *et al.*, 1990).

Heritability

The heritability of a trait explains how much of the phenotypic variance of that trait in a population is due to the genetic variance within the population. Heritability does not require prior knowledge of the genes controlling the trait in question and so can be used to understand the evolutionary forces affecting a specific trait (Wilson *et al.*, 2010). Traits that are heritable are amenable to selection. Heritability is measured on a proportional scale, where a value of 1 means that there is complete genetic control of the trait, and a value of 0 that the observed phenotypic variance is completely due to non-genetic forces, for example environmental variation.

Estimating heritability is contextual and measurements made in one population should not be applied among populations as the amount of genetic or environmental differences in other populations may vary and therefore affect a trait's heritability. The environment and the genotypic composition of a population may also vary over time, which can therefore also affect the temporal heritability of a trait. Studies of the heritability of the same trait in different populations often report different values of heritability (reviewed in Wells and Stock, 2011). Also, heritability is a population-level measure, and this does not give precise information about individuals in the population. For example, a trait with a heritability of 0.8 does not mean that an individual's phenotype is 80% determined by their genetics. Rather, 80% of the variance in the trait in the population is due to genetic variation.

As previously mentioned, heritable traits are amenable to selection, be it natural or artificial, therefore identifying whether a trait is heritable or not is of significant interest to animal breeders to assist in breeding the correct animals for specific traits

(Haskell *et al.*, 2014). Heritable traits also suggest that causal alleles, specific alleles that directly affect the phenotype of a trait, exist. With the advent of new sequencing and genotyping technologies, heritability studies can help guide the mapping of causal gene alleles through genome-wide association studies (GWAS) (Visscher *et al.*, 2008; Yang *et al.*, 2010; Thompson-Crispi *et al.*, 2014). GWAS studies look for statistically significant associations between loci and a specific quantitative trait and have stringent statistical filters to exclude loci with small effects on the trait. In doing this the identified causal loci in GWAS studies generally only explain a small proportion of the total heritability of complex traits, for example human height, and the heritability that is unaccounted is deemed the "missing heritability" (Visscher *et al.*, 2008; Yang *et al.*, 2010). New statistical approaches in GWAS studies, that include very large number of SNP loci simultaneously to calculate heritability, are now able account for much more of this "missing heritability" (Yang *et al.*, 2010). This resolution of "missing heritability" suggests that much of the heritability for complex traits, such as measures of immune function, may be due to multiple loci each with small effects on the overall trait phenotype.

Calculating Heritability

The phenotypic variance of a trait is the sum of the genetic and phenotypic variance in population as,

$P_{\text{(Phenotypic Variance)}} = G_{\text{(Genetic Variance)}} + E_{\text{(Environmental Variance)}}$

In its simplest form the heritability of a trait can be calculated by dividing the genetic variance in a population by the total phenotypic variance:

$H2_{\text{(Heritability)}} = V_{G\text{ (Genetic Variance)}} / V_{P\text{ (Total Phenotypic Variance)}}$

This calculation is known as the broad-sense heritability (*H2*) and includes all components of genetic variance: V_A additive genetic variation, V_D dominance variation, and V_I genetic interactions as,

$H2 = V_G (V_A + V_D + V_I) / V_P$

However V_D and V_I are difficult to measure and most studies focus on examining only additive genetic variance, which is the number of alleles shared by individuals (Wilson *et al.*, 2010). This is known as the narrow-sense heritability (*H2*) and is calculated by,

$H2 = V_A / V_P$

This compares the phenotypic similarities of differently related individuals to determine how strongly correlated phenotypic and genotypic similarity are, and therefore how much influence genetic variance has on the phenotypic variance of a trait. In the work presented here it is narrow-sense heritability, *H2* that will be considered.

The covariance between related individuals for quantitative traits is the basis for estimating heritability in classical quantitative genetic studies (Lynch and Walsh, 1998). To calculate narrow-sense heritability it is necessary to estimate both the total phenotypic variance of a trait, and genetic variance in the population. For a trait to be heritable it is necessary for genetic variance and different levels of genetic relatedness among the individuals to exist within a population. Pairwise genetic similarity is proxy of relatedness that is often used in the calculation of heritability. This is the average frequency among individuals that two homologous alleles, one in each individual, are identical-by-descent (Ritland, 2000). In other words that the allele has been inherited by both individuals from a common ancestor. Pedigree-based relationships of

individuals can act as a proxy for actual measures of relatedness. Parents pass on 50% of their genome to their offspring and therefore 50% of the offspring's genome is identical-by-descent to their parents. On average full sibling's genomes are also 50% identical-by-descent, however half siblings are on average only 25% genetically identical-by-decent. As previously mentioned, the covariance of genetic relatedness and phenotypic variance can give us the heritability of a specific trait. In humans the pairwise relatedness is relatively easy to calculate due to the recruitment of monozygotic and dizygotic twins in heritability studies or the use of known family pedigrees. In most wild populations of organisms, it is not possible to know the genetic relationships of individuals and therefore other proxies of genetic relatedness need to be calculated. The following sections describes commonly used study designs to calculate heritability among differently related individuals.

The Twin Approach

Estimating Heritability

Twin-based studies use the unique genetic concordance within, and discordance between, sets of monozygotic (MZ) and dizygotic (DZ) twins as a measure of genetic variation to calculate the heritability of traits. Complete identity-by-descent exists for sets of MZ twins compared to only 50% in DZ sets of twins. Twins can be useful in research since sets of twins also generally share identical environments, including *in utero* conditions. Sets of MZ twins derive from a single fertilised embryo and therefore they are genetically identical; somatic mutations may develop or differ between twins and can continue to be acquired following embryo separation. Epigenetic variation (dynamic changes in the regulation of gene expression) between twins will accumulate over their lifetime (Fraga *et al.*, 2005), however, this is largely not considered in heritability studies. In contrast, sets of DZ twins originate from two distinct embryos and, as with all full-siblings, are on average just 50% genetically identical to each other. Therefore, the heritability of a trait can be estimated by comparing the genotypic variance to phenotypic variance correlations between MZ and DZ twins. Narrow sense heritability can then be calculated as twice the difference between phenotypic correlations for MZ and DZ twins (Visscher *et al.*, 2008).

Twin studies have been used extensively to investigate the heritability of traits in human populations where MZ and DZ twins can be readily recruited. These studies have proved extremely valuable in identifying traits that are strongly influenced by genetic differences with over 15,000 traits being investigated in 2,748 studies (Polderman *et al.*, 2015). However, despite their genetic power, twin studies are limited in their application to wild populations. Firstly, they cannot generally be used in wild populations, unless detailed pedigree records are kept, since identifying MZ and DZ twins, which are rare, genetically is difficult. Secondly, in wild populations twins may experience considerably different environments from one another (for example due to competition, social interactions and resource limitations) much more so compared with humans where twins are often raised in identical conditions.

Recently it has become apparent that despite identical nuclear genomes MZ twins differ substantially in their epigenome; DNA methylation and histone modification, profiles (Fraga *et al.*, 2005). The epigenetic state of DNA is dynamic (Castillo-Fernandez *et al.*, 2014), with modifications occurring throughout life and are potentially reversible. MZ twins are often epigenetically indistinguishable at birth but with age epigenetic differences accumulate, causing significant epigenetic discordance within a MZ twin pair (Fraga *et al.*, 2005). Epigenetic modifications can affect gene expression, modify phenotypes and potentially affect the heritability of traits. If, due to epigenetic effects, MZ twins are genetically less concordant with one another than expected, heritability may be overestimated as the difference between MZ and DZ sets of twins would be reduced. A reduction in the estimate of a trait's heritability with increasing age, such as has been observed for various measures of immune function (Brodin *et al.*, 2015), is consistent with age-acquired epigenetic differences between twins. This has been demonstrated for many human traits including height, mental health disorders, endocrine function and blood pressure (Polderman *et al.*, 2015). Epigenetic differences increase with age as twins begin to be exposed to different environmental factors, including exposure to different infections. For example, discordance in past infection with cytomegalovirus infection in MZ twins results in greatly reduced correlations among twins for many immunological traits

such as $CD8^+$ T cell counts and serum concentrations of IL-10 and IL-6 cytokines, therefore lowering the estimate of heritability of these traits (Brodin *et al.*, 2015).

Human Twin-Based Studies of the Heritability of Measures of Immune Function

Numerous studies have examined the heritability of measures of immune function in humans with approximately one hundred studies conducted to date (Polderman *et al.*, 2015). Many studies examining the heritability of immunological traits fail to directly measure and estimate the heritability of measures of immune function, preferring mainly to examine the resistance/susceptibility of individuals to infection or disease. Examples of this are studies of psoriasis and eosinophilia esophagitis (EoE). These diseases are chronic, immunologically-mediated autoimmune diseases in humans. Twin studies have been used to investigate the heritability of these diseases and have found that neither are particularly heritable with values of only 0.34 and 0.15 for psoriasis and EoE, respectively (Grjibovski *et al.*, 2007; Alexander *et al.*, 2014). The studies did not investigate the heritability of the immunological traits associated with the disease.

For the studies that have directly investigated the heritability of measures of immune function the general consensus is that most traits are heritable, but the degree of heritability varies greatly among specific traits (Brodin *et al.*, 2015). For example, in humans one study reported the circulating serum concentration of cytokine IL-12p40 as highly heritable with a heritability ~1.0, whereas the circulating serum concentration of cytokine IL-10 had a heritability of just ~0.25 (Brodin *et al.*, 2015) However, in another study the heritability of IL-10 was estimated to be 0.62 (de Craen *et al.*, 2005). As discussed previously, the heritability of specific traits can vary from population to population and even on repeated studies of the same population (Wells and Stock, 2011).

Genetic variation among individuals can affect their risk of developing autoimmune diseases, and can also affect their responses to infection, vaccines and medication (Russi and Brown, 2015). Loss of balance between the cytokines produced by $CD4^+$ T-helper type 1 cells (Th1) and $CD4^+$ T-helper type 2 cells (Th2) has been suggested to

result in the development of chronic inflammatory conditions (Neurath *et al.*, 2002), therefore there has been considerable interest in determining the whether there is a genetic influence on Th1 and Th2 phenotypes. Holher *et al.* (2005) estimated the heritability of cytokines associated with each Th-cell phenotype using a cohort of German twin sets. This study found that Th1 cell cytokines were highly heritable, IFN-γ (0.85, 95% C.I. 0.74-0.95) and TNF-α (0.72, 95% C.I. 0.5-0.93) while the Th2 cell signature cytokine, IL-4, showed no significant heritability (Höhler *et al.*, 2005). These findings suggest that allergic and autoimmune diseases mediated by the imbalance between Th1 and Th2 cells are largely driven by genetic factors influencing the phenotype of Th1 cytokines, rather than that of Th2 cells (Höhler *et al.*, 2005).

De Craen *et al.* (2005) found that following *ex vivo* stimulation of heparinized whole blood with lipopolysaccharide (LPS) the *ex vivo* concentrations for the cytokines IL-1b, IL-1ra, IL-10, IL-6 and TNF-α had a heritability that was moderate to high, ranging between 0.53 – 0.86, suggesting that this cytokine production is under strong genetic control (de Craen *et al.*, 2005). Another twin study, examining a total of 204 immunological traits, found that circulating serum cytokine concentrations of IFN-γ, IL-12p70 and IL-4 all were highly heritable (*H2*, > 0.5), while IL-12p40 and IL-6 had heritabilities of ~1.0 (Brodin *et al.*, 2015). In contrast the heritability of IL-1a was below the detectable limit of the study, <0.2, and IL-1b had a heritability of only 0.25. The cell counts examined in this study showed much lower heritabilities with the total cell count for $CD4^+$ T cells, NK cells, monocytes, and B cells having heritabilities that were undetectable in the analysis. $CD8^+$ T cells had a moderate heritability of 0.3 (Brodin *et al.*, 2015). In total, of the 204 immunological traits examined in this study, only 23% were highly heritable and the authors suggest that for most immunological traits non-heritable factors such as age, infection history and vaccinations have the biggest effects (Brodin *et al.*, 2015). Age is known to decrease the heritability of numerous non-immunological traits in humans (Polderman *et al.*, 2015). Although MZ twins begin life having identical nuclear genomes mutations, epigenetic differences can occur independently in each twin and both accumulate with age. Therefore, over time MZ twins are no longer as genetically identical as they once were resulting in lower correlation of genotype to phenotype in older MZ twin sets; for example, the

heritability of immunological traits decreases from 0.7 in 12-17-year-old MZ twins to 0.56 in 18-64 year old MZ twins, and 0.4 in 65+ year old twins (Polderman *et al.*, 2015).

Pedigree Approach

Estimating Heritability

In the absence of twins, a pedigree of the population in question can be used to infer the genetic relatedness, via identity-by-descent, from the familial relationships among individuals. This genetic relatedness can then be used as the measure of genetic variation within the population. In calculating the heritability of a trait in this way the phenotypic similarity among individuals can be regressed onto the individuals' familial relationships, and the heritability calculated (Ritland, 2000). Calculating the relatedness of individuals using this method gives an associated error to the relatedness estimate, something which is not factored into twin-based studies (Ritland, 2000).

Pedigree-based estimates of heritability can be used with wild populations, but they do require a detailed knowledge of the population structure and genetic relationships among individuals. Therefore, these approaches are most often used with domesticated species where this information can be easily obtained, though these methods have also been applied to insects and nesting birds (Janss and Bolder, 2000; Gauly and Erhardt, 2001; Reiner *et al.*, 2007). While pedigree-based studies estimating heritability are an important approach they cannot be applied to novel populations, where pedigree information is not available. Pedigree studies are on average eighty-fold less efficient at capturing the heritability of a trait compared to twin studies; however they are still superior in their statistical power compared to marker-based approaches (Visscher *et al.*, 2008). In addition, if inbreeding may be underestimated by using inferences from the relationships of individuals, overall resulting in an overestimation of heritability.

Pedigree-Based Estimates of the Heritability of Measures of Immune Function

Pedigree-based studies of heritability require an extended pedigree of the study population, which includes multiple levels of familial relationship. Pedigree-based studies can be used with humans when familial records are sufficient enough to construct a pedigree. A pedigree study of 367 people from the Jirel population in Nepal was used to determine the heritability of immunological traits (Williams-Blangero *et al.*, 2004). The serum concentrations of six cytokines were found to be moderately (0.3-0.5) to highly (>0.5) heritable: IFN-γ (0.65, s.e 0.096), TNF-α (0.46, s.e 0.10), IL-2 (0.58, s.e 0.10), IL-4 (0.7, s.e 0.10), IL-5 (0.68, s.e 0.09) and IL-10 (0.60, s.e 0.09). The Jirael population is heavily parasitized by the helminths *Ascaris lumbricoides*, *Necator americanus*, *Ancylostoma duodenale*, *Trichuris trichiura* and the protozoan *Giardia lamblia* (Williams-Blangero *et al.*, 2004). The occurrence of these parasitic infections, as well as the sex and age of individuals, and the presence of a latrine in the home were all fitted as covariates in the heritability analyses to accommodate for the contribution to the trait variance of these factors (Williams-Blangero *et al.*, 2004).

Pedigree records of domesticated sheep, chickens and pigs are readily available, and parasitic infections of these species are major sources of productivity loss. Therefore, numerous studies have been conducted to investigate the heritability of resistance and immunity to infections in these species. In Rohn and Merionland sheep, parasite naïve lambs were experimentally infected with 5000 L_3 (larval stage 3) larvae of *Haemonchus contortus,* and the heritability of anti-*H. contortus* L3-specific IgL antibody titres (IgL antibody specific to *H. contortus* L_3 larvae) was estimated to be 0.13 and 0.3 in these breeds, respectively (Gauly and Erhardt, 2001). However, the IgG response to the same antigen had a heritability of 0 (Gauly and Erhardt, 2001)suggesting that neither are good candidates for selectable markers for resistance to *H. contortus* infection. In contrast in Pietran and Meishan pigs experimentally orally infected with 50,000 sporocysts of *Sarcosystis miesheriana*, there was a high heritability (0.74) for IgM antibody responses to the infection, and a moderate heritability (0.42) for anti-sporocyst-specific IgG responses (Reiner *et al.*, 2007). Another study of healthy domesticated Large White pigs examined the heritability of

total white blood cell (WBC) counts and the proportions of peripheral blood mononuclear leukocyte (PBML) cells positive for the cell surface markers: $CD4^+$, $CD8^+$, □□ T-cell receptor, CD11R1, B cell and monocyte markers, at the start and end of standard growth performance tests (Clapperton *et al.*, 2008). The authors found that the heritability for all these traits, at both time points were moderate to high (0.18 - 0.62) except for the proportion of $CD8□^+$ cells whose heritability was 0.18 (s.e. 0.13) (Clapperton *et al.*, 2008). This is in notable contrast to a human twin-based study where the total $CD8^+$ cell count, not proportion, was one of the only significantly heritable cellular immunological traits (Brodin *et al.*, 2015). In pigs at the end of the standard growth performance tests the heritability for WBC was 0.18 (s.e 0.11) and the heritability of the proportions of PBMLs that were different cell populations ranged from 0.44 (s.e. 0.14) for CD11R1+ cells to 0.62 (s.e. 0.14) for $CD4^+$ cells. Another study examining the heritability of immunological traits in Large White pigs, 54 immunological traits from heparinized blood samples stimulated with various antigens in pigs vaccinated with *Mycoplasma hyopneumoniae,* reported high heritabilities for the majority of immunological traits examined (Flori *et al.*, 2011). The immunological traits measured covered innate and adaptive, cellular and humoral parts of the immune system including WBC counts, cytokine serum titres (including IL4 and IFN-γ) and total antibody titres (IgG, IgA, IgM). In total 48 of the 54 immunological traits studied were significantly heritable with 18 being moderate (0.1-0.4) including IL-10 (0.35, s.e. 0.19) and 30 highly heritable (>0.4), for example IL-2 response to LPS stimulation (0.91, s.e. 0.19) (Flori *et al.*, 2011). Altogether, these studies principally of livestock, show that immunological traits can have a range of heritabilities, and in some settings heritability can be particularly high.

In wild populations extended pedigrees are rare, however the Soay sheep on the island of Hirta in the St. Kilda archipelago, Scotland, have been studied for over 20 years and a detailed pedigree, as well as parasitological and some immunological data, exist. Nematode infections in this population have been proposed to be an important selective force (Smith *et al.*, 1999) and locus-specific associations with measures of nematode parasitism have been identified (Gulland *et al.*, 1993), which suggested that resistance to these parasites is heritable. This has been tested in this population

through a pedigree-based study, which used molecular markers to confirm the parentage of individual sheep (Smith *et al.*, 1999). This study found that residual faecal egg counts, the remaining variation after a general linear model of non-genetic factors was fitted and is a measure of parasite resistance, was heritable (0.056 to 0.599) (Smith *et al.*, 1999). Another study of the Soay sheep sought correlations between the concentration of anti-nuclear antibodies (ANA), a proxy for overall antibody responsiveness, and measures of fitness such as overwinter survival rates (Graham *et al.*, 2010). This heterogeneity of ANAs was found to be significantly heritable (h^2=0.13, p <0.001) under an additive genetic model. Combined with the positive correlation between high ANAs and overwinter survival rate, this result suggests that immune function directly influences the fitness of individuals in the population (Graham *et al.,* 2010).

Insects have relatively simple immune systems, compared to mammals, however there is considerable homology between insect immune systems and elements of the mammalian innate immune system. Phenoloxidase and abaecin are two important antimicrobial molecules found in insect immune systems providing resistance to parasitic, bacterial and viral infections (Cotter and Wilson, 2002; Decanini *et al.*, 2007). The heritability of the amount of phenoloxidase (PO) produced and transcript levels of abaecin has been estimated using sibling analyses in the cotton leafworm (*Spodoptera littoralis*) and the honeybee (*Apis melifera*), respectively. Both traits were found to be heritable, for abaecin *H2* 0.3-0.4, and for PO at *H2* = 0.69 (Cotter and Wilson, 2002; Decanini *et al.*, 2007). These results suggest that genetic variance is an important component of these immunological traits and potentially for immune function in general and may therefore be a selectable trait among insect species.

Parentage of offspring can be easily determined in nesting birds, allowing for simple parent-offspring regression analyses as well as cross-fostering experiments to be conducted, so that the heritability of immunological traits can be determined. In these studies the immunological traits studied are often tissue swelling and the number of cells recruited to a site following immunisation with a novel antigen, particularly the mitogen phytohaemagglutinin (PHA) (Sakaluk *et al.*, 2015). Results from such studies

suggest that the immune response to PHA is, at most, only partially heritable. For example a cross-fostering study of tree swallows, *Tachycineta bicolor*, found significant heritability of inflammation at point of immunisation, a measure of immune function, in only one of the three populations of birds examined. Mothers with higher immune responses tended to have offspring with higher immune responses, but this was only found in a Tennessee population of birds, and not in the Alaska or New York populations (Ardia and Rice, 2006). The authors of this study suggest that temporal and spatial variation may play a significant role in the immune responses to PHA in these animals as only the Tennessee population studied had additive genetic variance, and therefore heritability of immune function, compared to the New York and Alaska populations. Similarly a non-sibling analysis of the collared flycatcher (*Ficedulla abicollis*) did not find significant heritability of the immune response to PHA, and this implies that sibling analyses may overestimate *H2* values (Pitala *et al.*, 2007). Kinnard and Westneat (2009) examined the heritability of body condition, growth and immune response (specifically T-cell mediated swelling, following PHA stimulation) in a partial cross-fostering study of the house sparrow, *Passer domesticus*. Condition and growth showed non-significant heritabilities, however the immune response was heritable (Kinnard and Westneat, 2009). Together these studies show that the heritability of the immunological trait of PHA response varies among species and even among different populations within the same study.

Marker-Based Approach to Estimating Heritability

Classical quantitative genetic estimates of heritability require an extended pedigree (Lynch and Walsh, 1998). However, in most wild populations these data are not available, and so the problem of estimating the genetic relatedness of individuals remains. An alternative approach to estimating genetic relatedness is to infer this through measuring the extent to which genetic markers are shared among individuals (Queller and Goodnight, 1989; Ritland, 1996; Goodnight and Queller, 1999). A positive correlation between relatedness, inferred from the number of shared markers among individuals, to phenotypic similarity suggests that a trait is influenced by genetic variation (Ritland, 1989). Improved molecular techniques and technologies have allowed for genetic markers to be more effective as measurements of relatedness and of population structures, than they had been previously (Cruzan, 1998). This is due to a greater coverage of markers, and a wider variety of statistical tools which can calculate inferred relatedness. Even in pedigree-based studies, molecular markers are often used to estimate or confirm the parentage of individuals (Hughes, 1998; Jones and Ardren, 2003). However, there have been relatively few studies that have used a marker-based approach to calculate the heritability of quantitative traits (Ritland and Ritland, 1996; Klaper *et al.*, 2001; Wilson *et al.*, 2003; Yang *et al.*, 2010). Even fewer studies have applied marker-based estimates of relatedness to wild populations (Thomas *et al.*, 2002; Coltman, 2005; Kumar and Richardson, 2005). The results of these marker-based studies of heritability have been found to have poor accuracy and yield unreliable parameter estimates, compared with classical methods to calculate heritability (Coltman, 2005). This is possibly due to high sampling variance or a low amount of relatedness variance in a population (Van De Casteele *et al.*, 2001). However, new analytical methodologies and approaches applied to marker-based estimates of relatedness are promising, giving more reliable heritability estimates, that are more consistent with estimates calculated using classical approaches (Kumar and Richardson, 2005; Yang *et al.*, 2010; Gay *et al.*, 2013). Also, it is worth noting that when an extended pedigree of a population is not available, as with the majority of wild animal populations, marker-based estimates are the only tangible option for estimating the heritability of a trait in that population.

Caution must be applied to any results generated from marker-based studies of heritability, because of the ability of a marker-based study to capture the heritability of a trait is dependent on linkage disequilibrium occurring among the markers used and the causal loci underlying the trait in question. Therefore, in order to maximise the ability to identify LD between markers and causal loci a high density of markers is recommended (Visscher *et al.*, 2008). However, with respect to heritability studies of immunological traits in wild populations, until now the genetic and immunological resources have not been readily available for most non-model species. This has prevented marker-based heritability studies of measures of immune function in wild populations.

As discussed previously, our understanding of laboratory mice, particularly their immunology and immunogenetics is vast. In contrast, our knowledge of the immunology and immunogenetics of wild mice is substantially more limited, because of the limited study of house mice *per se*, but particularly limited study of their genetics and immunogenetics.

Laboratory mice differ from house mice in two critical ways. Firstly, lab mice have been selectively bred and inbred over many generations meaning that they (i) have limited genetic variation and (ii) are genetically and phenotypically diverged from wild mice. Secondly, lab mice live in research facilities that are generally pathogen free and where the animals are provided with a perfect diet *ab libitum*. In these two ways – both historical and contemporary – wild and lab mice are different. Therefore, the current unknown is the extent of the functional immunological difference between wild and laboratory mice. This is a key question because (i) the immune state of wild mice has hitherto not been studied *per se* and, (ii) that the immune state of wild and laboratory mice has rarely been compared. Current gaps in our knowledge are therefore, what is the immune state of wild mice and how does this differ from the immune state of lab mice.

Beyond the immune state of wild mice, the extent of genetic variation in genes coding for key molecules of the immune system is not known, because this has not been studied before. The immunogenetic variation in laboratory mice is due to selective breeding and inbreeding and it is not known whether this genetic variation in any way represents genetic variation that occurs in wild mice. Current gaps in our knowledge are therefore the nature and extent of immunogenetic variation in wild mice, and whether and how this differs from what is known from study of laboratory mice.

Aims

The aim of this study is to examine the genetic variation that exists in a wild population of house mice and to determine how this variation may influence measures of immune function in these mice. To do this I will sample mice from various locations in southern Britain and evaluate the genetic variation that exists among these mice and then associate this with variation in measures of immune function.

Hypotheses

- Strong genetic subdivision exists between mice geographically isolated from each other.

- The heritability of immunological traits will be low due to the power of the study and the polygenic nature of the traits investigated.

- Considerably more genetic variation will exist in genes of the mouse H2 locus, due to the selective pressures of many variable pathogenic infections.

- The presence of any non-synonymous SNPs will have significant effects on measures of immune function in comparison to synonymous or non-coding SNPs.

Chapter 2: The Population Genetics of Wild House Mice, Mus musculus domesticus

Introduction

Abstract:

Most populations of house mice are commensal, found in close proximity to humans, and therefore the genetic makeup of these mice is greatly affected by human activities. These commensal house mice live in demes, small breeding units, where migration is rare and therefore gene flow is inhibited. Feral house mice are less dependent on human activities and more free breeding.

In order to determine the genetic population structure of wild house mice in Southern Britain over 400 house mice were genotyped using over 1000 neutral SNP loci. 12 populations were sampled in total, one from London Underground, one from one a feral population from Skokholm island and the remaining 10 from farm and equestrian locations around Bristol and Gloucestershire. Analyses of the genotypic data suggests that commensal and feral house mice have strikingly different breeding structures. Inbreeding occurs most frequently for the Skokholm Island, feral, population demonstrated by the strong clustering of these mice on a nearest-neighbour joining tree and supported by STRUCTURE analysis. F_{ST} values among all sampled sites were high and significant, suggesting substantial population subdivision with very little gene flow between these sub-populations. These findings suggest that the gene flow among commensal house mice is limited and that this is due to human activities such as farming practices and pest control.

Advances in our understanding of how genetics affects health and disease have largely come from the intensive study of the genetics of the house mouse, *Mus musculus domesticus*. This work has led to the identification of numerous loci that contribute to disease states, which has also led to the identification of therapeutic drug targets. Together this has significantly improved the treatments available for a large number of immunological diseases such as rheumatoid arthritis and type-1 diabetes (Morel, 2004). These genetic studies have used standard inbred laboratory strains of mice, which give the required analytical power to establish the relationship between genotype and disease phenotype.

In contrast to this very substantial research effort, the biology and genetics of wild house mice have received relatively little research attention. This is despite house mice being one of the most widely distributed species on the planet (Singleton and Krebs, 2006), and as such a significant pest species, causing significant economic losses (Brown and Singleton, 2002). Wild mice live in highly heterogeneous environments and, like other free-living mammals, their populations are assumed to contain genetic variation, in notable contrast to inbred, laboratory strains of house mice. Wild mice may therefore be an important reservoir of rare genetic and phenotypic variants absent in inbred laboratory house mice. This additional diversity could potentially be tapped to allow for new, fine-scale mapping studies of traits of biological and biomedical interest, and so possibly further improve and develop conventional laboratory studies. However, because so little is known about the genetics of wild mice, a key first step is to determine how much genetic variation exists in wild house mouse populations, and the nature of the population structure of such animals.

Research into house mice began at the start of the twentieth century (Castle and Allen, 1903; Phifer-Rixey and Nachman, 2015). Early studies involved the genetic mapping of loci controlling the coat variation of so-called “fancy mice” (Castle and Allen, 1903). Such forward genetic studies were rapidly expanded beyond study of “fancy mice”, when mice were further crossed and selectively bred for other phenotypes of interest from a single, female, mouse (Ferris *et al.* 1982). This process led to the development

of the highly characterised, inbred strains of mice that are today commercially available, and used extensively in research laboratories around the world.

Sequencing of the house mouse genome, C57BL/6 strain, was completed in 2002 (Waterson *et al,.* 2002). This then led to the advancement of the reverse genetic approaches with mice, via techniques such as gene knock-outs and site-directed mutagenesis (reviewed by Nguyen & Xu 2008). In comparison to forward genetic studies, reverse genetic studies seek to determine the function of a gene through observations of phenotypes (Nguyen and Xu, 2008). This approach has allowed the identification of various disease-associated loci, which has then been applied to human biology. Transgenic mice, mice carrying genes originating from another species, have been developed more recently (Gordon and Ruddle, 1981). These mice allow for more detailed study of the aetiology of diseases and the effects of drugs. These sophisticated genetic manipulations have only been possible due to our detailed genetic knowledge of the humble house mouse. Continued research into the genetics of house mice has also led to the development of genetic and genomic techniques and resources unique to the species. However, until now the application of these tools has been largely restricted to laboratory house mice. The few studies that have applied these sophisticated tools to wild house mice have primarily focused on the phylogeny and macro-population structure of sub-species of house mice (Salcedo *et al,.* 2007).

Population Genetics of *Mus musculus domesticus*

The majority of population genetic studies of *Mus musculus domesticus* were conducted over 40 years ago using allozymes and microsatellites (Berry & Jakobson 1974; Baker 2016; Anderson 1964; Petras 2016a; Selander 1970). These studies fall mainly into two categories: studies of commensal house mice and of feral mice. Commensal mice live in close proximity with humans and human habitation and their lifestyle and habitats are dependent on human activity. In contrast feral house mice live independently from humans in more open, natural environments. The differences between commensal and feral mice in their respective environments and lifestyles have important effects on the mouse population structure and gene flow that occurs among sub-populations.

Commensal House Mice

Studies of commensal house mice largely show that these mice exhibit high genetic sub-division among populations. The cause of this population sub-division is likely due to the isolated breeding structure of *M. musculus domesticus* (Anderson *et al.,* 1964; Anderson 1964; Crowcroft & Rowe 1963; Reimer & Petras 1967; Lewontin 1962; Lewontin & Dunn 1960; Petras 1967a; Petras 1967b), discussed below. Until recently variation in allozymes, in particular the haemoglobin (Hb) and Esterase, (Es-1 and Es-2) loci (Petras, 1967a; Selander, 1970; Baker, 1981), have been some of the primary investigatory tools used in population genetic studies of wild house mice. Petras (1967a) investigated the allele frequencies of the Hb, Es-1 and Es-2 loci in commensal mice living in different buildings within a single farm, and among farms and in different years. Focusing on the Es-2 locus the study found that F_{ST} values fluctuated between years, but overall the F_{ST} value among mice from all sampling locations and all years was 0.18, meaning that 18% of the genetic variation among the mice was attributable to differences among sub-populations. This is a moderate value of F_{ST} for a mammalian species (Petras 1967a). The highest F_{ST} value reported was 0.441 for mice inhabiting an isolated barn on one farm. This value is extremely high for a mammalian population, indicating very high genetic separation of these mice from the wider population. Further examination of F_{ST} values among farm buildings and among farmsteads showed that the majority of the population sub-division occurred on a fine

scale, specifically among different farm buildings, with an average value of 0.172. In contrast, relatively little genetic sub-division was attributable to differences among mice from different farmsteads. Overall, this study suggests that commensal mice can become genetically isolated across relatively small geographical distances. The allele frequencies and inbreeding co-efficient of the Hb and Es-1 loci are also consistent with the idea that commensal mice live in small isolated breeding units known as demes, which is discussed further below. However, it is also worth noting that as these results are drawn from the analysis of single loci, they may not necessarily give a true representation of the genetic differentiation of commensal house mice at other loci, or in fact the entire genome, and therefore further investigation is required.

Selander (1970) also utilised allozyme data, for the haemoglobin Hbb locus (*Hb)* and esterase 2 and 3 (*Es)* loci, in the study of wild house mice inhabiting farms across Texas. Commensal mice inhabiting the same farm, or even building, demonstrated spatial population genetic heterogeneity, with greater genetic similarity occurring among individuals that live in close proximity. The author refers to this as "tribal subdivision" (Selander 1970). Therefore, this study is consistent with the findings of Petras (1967a) in providing evidence of genetic sub-division among commensal house mice even across relatively small distances.

The recessive lethal T/t locus (chromosome 17) has also been used in the study of population genetics of wild house mice. The *t*-allele is recessive-lethal, however meiotic drive increases the transmission of the *t*-allele to offspring of heterozygote males to 96% (Berry and Jakobson, 1974). Anderson (1964) investigated the differences in *t*-allele frequencies among mice inhabiting grain storage buildings on Canadian farms, with each granary located well within the dispersal range of house mice (Pocock *et al.,* 2005). Anderson (1964) found that the differences in the allele frequencies of this locus persisted over several generations and that the mice in each granary were an isolated breeding unit, or deme. When *t*-allele carrying mice were introduced to a group of mice where the *t*-allele was initially absent, the introduced allele spread quickly among the closest mice. However, the dispersal of the allele into the more physically distant populations was slower (Anderson *et al.*, 1964). This is

further evidence that within these house mouse populations mating occurs predominantly among individuals that are in close proximity to one another. Bennet *et al.* (1967) explored this further by comparing the spread of the *t*-allele, introduced to a population on Great Gull Island, to theoretical predictions of *t*-allele frequencies from computer-generated models. It was found that the rate of propagation of the *t*-allele was slower than would be expected in a single randomly breeding population (Bennett *et al.*, 1967). The optimal breeding structure to achieve the observed t-allele frequency in wild house mice, 0.3, has been tested experimentally. A small breeding unit with few reproductive individuals, two males and six females, was identified as the most likely breeding structure to achieve this allele frequency (Lewontin and Dunn, 1960; Lewontin, 1962). In contrast to these studies Baker (1981) demonstrated that alleles newly introduced to a population are capable of spreading among individuals inhabiting different farm buildings in only a few generations. This suggests that the animals' breeding structure and distance between farm buildings are not a barrier to gene flow. However, this study primarily examined gene flow across a relatively small scale (<100m) of distance.

Most population genetic studies of wild house mice show that migration, and therefore gene flow, among sub-populations is restricted to some degree across distances of less than 100m. The reduced migration among sub-populations results in the allele frequencies of these sub-populations diverging over time due to genetic drift. Drift progresses more rapidly in small populations and has significantly contributed to the genetic differentiation observed among commensal house mice, even those in relatively close proximity to one another (Anderson, 1964; Petras, 1967a). It is therefore reasonable to assume that across greater geographical distances migration becomes even less frequent and genetic separation among sub-populations becomes greater. Increased frequency of mating between individuals in close proximity suggests that isolation-by-distance (Wright 1946) will affect the genetic flow within a population, creating significant genetic subdivision. However, isolation-by distance among wild house mice is yet to be studied using modern molecular markers.

Behavioural studies also show that wild, commensal house mice live in small territorial breeding units called demes. A deme is formed of a dominant male, breeding females and their offspring; subordinate males may occasionally contribute to a deme (Reimer and Petras, 1967). Because of this deme-based breeding DeFries and McClean (1972) suggest that the effective size of a breeding population may be as small as four individuals (DeFries and McClearn, 1972). Both sexes actively defend territories, however only males establish a stable territory (Reimer and Petras, 1967). The migration of foreign mice among demes is rarely successful with 91% of migrant males and 72% of migrant females killed by residents (Reimer and Petras, 1967). In experiments where mice were transplanted between granaries on a Canadian farm none of the transplanted mice were found inside the granary when resampled, with several having migrated back to their original granary (Anderson 1964). The rigid breeding structure of commensal house mice clearly poses a potential barrier to gene flow, resulting in genetic differentiation and the formation of sub-populations.

Feral House Mice

The habitats and behaviour of feral house mice differ substantially from commensal mice. Feral mice do not have the rigid breeding structure and territoriality of commensal house mice (Reimer and Petras, 1967). The habitats of feral house mice are often resource and shelter poor and as a result feral mice are less aggressive and more co-operative in their behaviour than their commensal counterparts (Jones *et al.,* 1995). One particularly well-studied population of feral house mice is located on Skokholm Island, off the coast of Wales. The island's mouse population was founded in the mid 1880's (Berry, 1964) and has been well studied since the 1960's. The mice of Skokholm Island are well dispersed across the island and experience high winter mortality rates when breeding also ceases (Berry and Jakobson, 1974); in contrast commensal house mice breed all year round (Pocock *et al.*, 2004). The seasonal mortality results in repeated population bottleneck events (Berry and Jakobson, 1974). Although less intense than is observed among commensal house mice, territoriality has been observed on the island, particularly among male mice (Berry and Jakobson, 1974). At the start of a new breeding season male territories have some degree of overlap, but these become more distinct throughout the year (Berry and

Jakobson, 1974). Females generally move into the territory of a single male (Berry and Jakobson, 1974). The turnover of dominant male territory is much higher than that observed among commensal house mice; in one breeding season three successive males have been observed to successively claim a single territory (Berry and Jakobson, 1974). Together, these findings suggest that subordinate males make a greater contribution to the gene pool of a population in feral house mice compared to commensal house mice. Jones *et al.* (1995) transplanted mice from a commensal population of house mice (from the Isle of Eday), to a feral island population (on the Isle of May). When this was done mtDNA and Y-chromosome DNA markers spread from the introduced commensal mice among the resident, feral mice (Jones *et al.*, 1995). While both markers spread through the population, the Y-chromosome marker did so at a much greater rate. This study suggests that while in commensal populations of house mice male mice are more likely to be attacked and less likely to integrate into an existing (native) deme, in feral populations foreign males are preferentially mated with. This preference for foreign males may be due to females actively avoiding inbreeding through mate choice or come about by the foreign mice out-competing resident mice. However, as the introduced male mice originate from a commensal population, they may be inherently more aggressive and outcompete the native feral males and therefore may not reflect a foreign male preference but merely commensal male dominance to feral males.

Migration rates of house mice among demes on Skokholm Island have been estimated at 30% per generation from a mark release recapture study (Berry and Jakobson, 1974). This is a substantial increase in migrants compared to that observed in commensal house mice, where only 1-5% of individuals were observed to be migrating between buildings on a single farm per generation (Petras, 1967a; Baker, 1981). The more frequent migration between demes and the consequent higher turnover of males in a deme in feral mice, means that although in feral mice the size of a single breeding unit may be equivalent to that commensal populations, the effective size of a breeding population of feral mice is much higher (Berry and Jakobson, 1974).

Large Scale Multi-Locus Studies of Wild Mice

More recently population genetic studies of mice have used microsatellite- and mtDNA-based data to investigate the ancestral origins of different house mouse populations. Searle *et al.* (2009) used restriction fragment length polymorphisms (RFLP) of nuclear markers and mtDNA sequences to compare mice sampled from 105 different localities in Britain (Searle *et al.*, 2009). This study found that genotypes identified among house mice from the Orkney Islands were highly similar to those of Norwegian mice. These results suggested that the shared ancestry of these two distantly located populations might be the result of historic human movements, specifically the migration of Vikings from Scandinavia to the present-day UK. The commensal nature of most house mice means that human migration and transport has played an important role in shaping the population structures of wild house mice. A study examining the genetic sub-division among mice inhabiting nine islands of the Azores archipelago, the island of Madeira and mainland Portugal highlights this (Gabriel *et al.*, 2013). Using 19 microsatellite markers the authors found that there was strong genetic sub-division among these mouse populations. This genetic sub-division was observed in a hierarchical manner with pairwise F_{ST} values greatest among the Western, Eastern and Central groups of islands, then among individual islands within these groups. The degree of observed genetic sub-division among islands and the mainland strongly reflects the human movements and particularly shipping routes (Gabriel *et al,.* 2013).

Aims:

- Determine how much genetic variation exists among these mice?
- Examine the effect of geographical distance on genetic isolation
- Identify the population structure of wild house mice in Southern Britain

Materials and Methods

Study Sites, Trapping and Culling

Wild house mice were trapped at 12 different sample sites in southern Britain (Figure 2.1) between February 2012 and April 2014 using Longworth live-traps. Henceforth all sample sites will be referred to using the site codes shown in Figure 2.1. The majority of sample sites were farms surrounding Bristol and in Gloucestershire. Two out-groups were also sampled, specifically mice from London Underground (LU) and Skokholm Island (SK). A total of 460 wild mice were caught over this period.

Trapped mice were transferred to the University of Bristol and temporarily housed in standard laboratory mouse housing. Each individual was assigned a numerical identification reflecting their order of capture. The weight and sex of the animals were recorded at time of capture. Animals were killed within two weeks of capture. The culling and processing of these mice is described in Weldon *et al.* (2015) and carried out by me, Stephen Abolins and Laura Weldon.

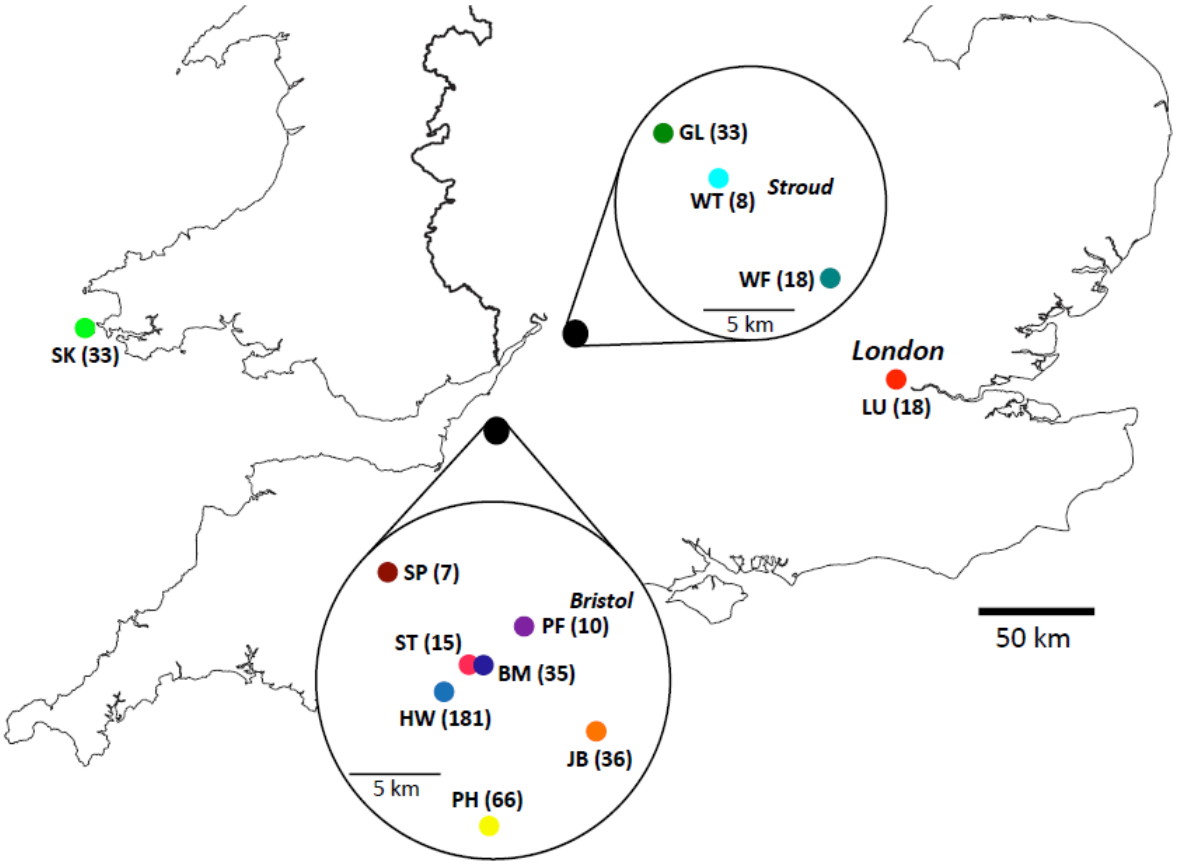


Figure 2.1

A map of the sample locations used in the current study in Southern Britain, adapted from Riley *et al.* in prep. Sites are identified by their two letter codes and colour coded, and the number of mice sampled at each site in brackets. BM, Barrow Mill; GL, Green Lanes Farm; HW, Hyatts Wood Farm; JB, J.B Equestrian; LU, London Underground; PF, Parsonage Farm; PH, Pixie Hall Farm; SK, Skokholm Island; SP Sperrings Farm; ST, St. Katherine's Farm; WF, Woefuldane Farm; WT, Westrip Farm. Adapted from Viney *et al.* (2016)

Site	**BM**	**GL**	**HW**	**JB**	**LU**	**PF**	**PH**	**SK**	**SP**	**ST**	**WF**	**WT**	**Total**
No. Mice	33	30	167	36	18	10	63	30	6	15	18	7	433

Table 2.1. A summary of the number of individuals successfully genotyped from each of the studied sample sites

DNA Extraction

Genomic DNA was extracted from the tail tips of the wild house mice. 5mm of tail tissue for each mouse was digested with 20μl proteinase K (Thermoscientific, UK) in 480μl digestion buffer (10mM EDTA, 1% w/v SDS, 30mM Tris.Cl pH 8.0) for 12hrs at 56^{o}C. Following digestion 20μl of RNAseA 10ng/μl (Thermoscientific, UK) was added to the samples and incubated at room temperature for 10 minutes. 700μl of phenol (pH 8.0) was then added to each sample and thoroughly mixed; this was then centrifuged at 16,000 g for 2 minutes. The top, aqueous, layer was then moved into a new micro-centrifuge tube with 500μl of phenol: chloroform: isopropyl alcohol (pH 8.0) (Thermoscientific, UK) in the ratio of 25:24:1, and mixed, this was then centrifuged, as above. The top layer was then moved into a new micro-centrifuge tube with 500μl of chloroform and mixed; this was followed by a further centrifugation, as above as above. The top layer was again aspirated into a micro-centrifuge tube and 980μl of 100% ethanol and 20μl of 3M sodium acetate was added. This was then centrifuged again, as above, and the ethanol was aspirated from the tube and 70% v/v ethanol added to wash the DNA pellet. This was centrifuged for a final time, as above, and the ethanol solution aspirated off. The remaining DNA pellet was then air dried and resuspended in 100μl of TE buffer pH 8.0 (30mM, Tris 10mM EDTA).

A nanodrop spectrophotometer was used to determine the purity of DNA. The absorbance ratios of 260:230 nm wavelengths were used to check for chemical contaminants such as phenol and chloroform. The target range for this ratio was between 1.8 and 2.35. The absorbance ratio of 260:280nm wavelengths were used to check for protein contaminants of DNA samples, the target range was between 1.6 and 2.2. Where DNA samples were insufficiently pure they were cleaned by repeated phenol chloroform purification and a second ethanol precipitation, as described above. If samples failed for a second time they were discarded, and a new extraction conducted. DNA was then quantified using the Qubit fluorimeter (Thermoscientific, UK) using the broad range double stranded DNA kit (Thermoscientific, UK) according to the manufacturer's instructions. 2μl of extracted DNA was used in the quantification. DNA samples were then diluted as necessary for the various

downstream processes they would be used for. DNA samples in TE buffer were stored at 4°C for immediate use or -20°C for longer storage.

Golden Gate Genotyping

For Golden Gate genotyping (Illumina, UK) 250ng of DNA was provided for each mouse at a concentration of 50ng/μl in TE buffer. In total 445 mice were genotyped including 443 wild mice and two laboratory C57BL/6 mice as genotyping controls, designated as L88 and L90, which were female and male, respectively. DNA samples from both these control mice were replicated three times on each of the five genotyping plates that were used, resulting in a total of 15 repeats for these controls. These repeats were later used as quality controls for the genotyping of the wild mice.

The genotyping of these wild and control samples was performed by Wendy McArdle (School of Social and Community Medicine, University of Bristol). The genotyping was carried out according to the manufacturer's instructions using the Goldengate Mouse MD Linkage OPA (Illumina, UK), GGGT Universal-32 1536 plex Assay kit (Illumina, UK), Golden gate DNA activation kit (Illumina, UK) and Titanium Taq DNA polymerase (Clontech, France). Golden Gate genotyping uses allele-specific oligonucleotides (ASOs) to hybridise with template DNA. Locus-specific oligonucleotides (LSOs) are also required to allow for down-stream hybridisation of single stranded DNA, following PCR. These PCR products hybridise to specific locations on the genotyping chip and each site is read for a fluorescent signature, which is dependent on the ASO that originally hybridised to the template. The Mouse MD linkage assay genotypes 1449 loci across the mouse genome, covering 19 autosomes and the X-chromosomes. On average the SNP loci are distributed with three SNP loci per megabase (Mb) of sequence. The SNP loci used on the genotyping chip were selected based on their known to be polymorphic among laboratory house and are also non-coding and assumed to be neutral under selection. The data were checked against the internal quality controls before any further analysis. The control mice were then used to check for accuracy and reliability by comparing the genotypes of these mice to the C57BL/6 reference genotype and comparing repeated genotypes to one another. Genotypic

data for the same loci was also obtained from Illumina for 10 laboratory-inbred strains of house mice for comparison to the wild mice.

Genotype Analyses

Neutrality

There are several programs that test for the neutrality of loci; in the current study Bayescan 2011 (Foll and Gaggiotti, 2008) was used. Bayescan uses a multinomial-Dirichlet model of allele frequencies, in which the F_{ST} coefficient (see F_{ST} below) is deconstructed into population-specific (beta β) and locus-specific (alpha α) components. If alpha values of a locus are significantly different from zero this suggests that the locus is under selection, and therefore not neutral. Positive values of alpha suggest that the locus is under diversifying selection where heterozygotes are favourably selected, while negative values of alpha suggest balancing or purifying selection, where a single homozygote genotype is advantageous. Bayescan analyses are also able to tolerate uneven and small sample sizes. In the current study the neutrality of loci was analysed using 1183 loci, following the removal of loci with a low minor allele frequency, and 433 wild mice using Bayescan 2011 (Foll and Gaggiotti, 2008). The data were provided in the co-dominant SNP data format.

Arlequin

HWE, F_{ST} values and AMOVA analyses were carried out on the genotypic data from 433 mice from 12 sample sites (Figure 2.1) using 1168 loci, after the removal of non-neutral loci, using the program Arlequin 3.5 (Excoffier and Lischer, 2010). The genotype data of the wild and control mice were first converted to the Arlequin data format using the CONVERT programme (Glaubitz, 2004). For these analyses mice were analysed as a single population and each sample site was treated as a separate sub-population.

HWE was calculated for each locus and significant deviations from HWE were tested with the Markov chain exact test in Arlequin 3.5 (Excoffier and Lischer, 2010), with one million iterations. This analysis was carried out on all mice considered as a single population, and then by each sample site separately.

F_{ST} was calculated using all loci and all mice, and in pairwise comparisons between mice from different sample sites. Significance was tested using 1000 repeats. To account for the large disparity in sample size among sample sites a repeated measures analysis was carried out three times using 6 randomly selected individuals from each site.

AMOVA analysis was conducted for all mice and, then in a hierarchical manner with sample sites grouped as: (i) a three-group analysis, mice were grouped as LU, SK and the remaining Bristol and Gloucester sites; (ii) A four-group analysis included the LU and SK sites as separate groups, however the Bristol and Gloucester sites were divided with the Bristol group including BM, HW, PF, PH, SP, ST and JB and the Gloucester group GL, WF, WT (Figure 2.1)

Linkage Disequilibrium (LD)

Linkage disequilibrium was calculated using the program Haploview (Barrett *et al.,* 2005) using 1168 loci and 433 wild mice. Loci were analysed chromosome by chromosome. LD scores were recorded as D' values, which range from 0-1.

MEGA

Nearest-Neighbour Joining (NNJ) trees were constructed in MEGA 6 (Tamura *et al.,* 2013) based on data from 1168 loci for 433 wild mice. Genotype data for each mouse was converted into a FASTA format and then a pairwise distance matrix of the number of nucleotide distances between each individual sample was calculated, with missing data accounted for in a pairwise manner. A bootstrapped NNJ tree was then constructed, with 1,000 bootstraps.

STRUCTURE

STRUCTURE (Pritchard *et al.*, 2000) analysis was carried out to more finely resolve the population structure of the wild house mice. 1168 loci were used in the analysis under the admixture model in order to identify individuals with mixed ancestry among sub-populations. Analyses were carried out with K values of 1-12 in order to determine if

each separate sample site (Figure 2.1) was a distinct sub-population. 15 independent repeats were carried out for each value of K with 100,000 Markov Chain Monte Carlo iterations and a burn-in length of 50,000. Analyses were run under the correlated allele frequency model on the assumption that individuals have a common ancestry. Initial STRUCTURE analysis was conducted using all 433 mice. A repeated measures analysis was then carried out using 6 randomly selected mice from each sample site, with this repeated three times. In this analysis the same individuals as used in HWE, and F_{ST} repeated measures were used. This repeated measures analysis was done to correct for the disparity in sample sizes across sample sites, given that STRUCTURE results can be strongly biased by uneven sample sizes (Kalinowski, 2011). All settings of the repeated measures analyses were the same as the initial analysis, with only five independent repeats being carried out for each of the repeated measures analyses.

The most likely value of K was determined using the Structure Harvester web service (Earl and Bridgett, 2012). This program uses the Evanno method of posterior probability estimation (Evanno *et al.,* 2005). This was used on the initial STRUCTURE run results and then with the repeated measures analyses grouped together. The STRUCTURE runs were then compiled and plotted using CLUMPAK (Kopelman *et al.,* 2015).

Isolation-by Distance

Pairwise F_{ST} and geographical distances were analysed using a Mantel test of correlation in Excel, using the GenAlex add-in. Geographical distances were measured in metres among the plotted sites, as-the-crow-flies, in Google maps (Jensen *et al.*, 2005), log10 distance was used in analyses. Pairwise F_{ST} values were obtained from prior analyses in Arlequin 3.5 (Excoffier and Lischer, 2010). 99 permutations were used in the mantel test calculation.

Results

Genotyping and Data Cleaning

In total 443 wild house mice from 12 sample sites (Figure 2.1), and two control C57/BL6 mice were genotyped at 1449 loci. 10 mice were excluded from further analysis due to poor quality genotyping data, leaving 433 mice for analysis. 99 loci located on the X chromosome were excluded from further analyses, because in males these loci are hemizygous. 144 loci were monomorphic in all mice and therefore would be uninformative in any future analyses, and so these were also excluded from further analyses. 23 of the loci genotyped poorly, specifically missing genotype data for more than 5% of mice, and these loci were also excluded from further analysis. The final data set therefore contained 1183 loci and these data were used for all subsequent analyses. Table 2.1 shows the number of mice that were successfully genotyped at these loci.

Controls

Two laboratory C57BL/6 mice were used as genotyping controls, one male and one female. Each mouse was genotyped 15 times in total. C57BL/6 is a highly inbred strain of house mouse and as such individuals should have identical genotypes. Comparison of the repeat genotypes for each of these mice measured the total genotyping error rate and so the reliability of the genotype data more generally. Across all possible 2366 nucleotides (1183 diploid SNP loci genotyped) the overall error rate was 0.96%. Most errors occurred in just one repeat genotyping of mouse L90, and if the data for this repeat genotype are excluded, then the overall genotyping error rate was 0.36%. Overall, the low error rate from these control mice suggests that these data are highly reliable.

Neutrality

Bayescan analysis revealed that of the 1183 loci tested, 15 were non-neutral having a *p*-value of <0.005, and therefore predicted to be under selection, see Table 2.2. These loci were located on 10 of the 19 autosomes, with chromosomes 4, 7, 8, 10 and 15 each having two of these loci. Five of the non-neutral loci were located within annotated genes: rs13479477 (Trim30e-ps1), rs13480652 (Pcdh15), rs6256918 (Cpsf6), rs4230248 (Arhgef3) and rs13482744 (Krt18) (identified using IGV and the mmouse 10 genome). Two of these loci (rs13479477 and rs6256918) were located in the exons of genes.

Locus I.D	Chr	Genomic position	3' Flanking	3' Distance (bases)	5' Flanking	5' Distance (bases)
rs3713616	1	74900046	Cryba2	6903	Mir375	612
rs13476649	2	95929798	Api5	1491651	Lrrc4c	388371
rs3712541	4	59103430	Gng10	61531	Ugcg	86120
rs13477765	4	69553312	Brinp1	598915	Cdk5rap2	857055
rs13478801	6	65211036	Hpgds	66306	C130060K24Rik	247114
rs13479375	7	71406934	BG130024G19Rik	995788	Gm36633	540593
rs13479477	7	104533846	Trim30e-ps1		Exon 3	
rs6153168	8	8294541	Slc10a2	3189309	Efnb2	322898
rs13479922	8	91879489	Irx3	77835	Crnde	446542
rs13480652	10	74437771	Pcdh15		Intron 16-18	
rs6256918	10	117348939	Cpsf6		Exon10	
rs4230248	14	27218513	Arhgef3		Intron 3	
rs13482501	15	29311563	Dnah5	839518	Ctnnd2	861030
rs13482744	15	102030597	Krt18		Intron 3	
rs3671671	19	4208173	Rad9a	6570	Clcf1	6219

Table 2.2. The 15 loci identified to be non-neutral, showing their chromosomal location and the position and identity of the adjacent 5’ and 3’ genes (IGV mouse, mm10). Loci in boxes are within a known gene, exon or intron.

LD

For Haploview analysis, the D' score was used as the measure of linkage disequilibrium. D' scores over 0.75 were considered to indicate strong linkage among loci. The majority of loci showed very low amounts of linkage among them. In total only 16 loci exhibited a D' score above 0.75, (Table 2.3), suggesting that the mice in this study are highly outbred. Chromosome 5 had the greatest number of linked loci with three pairs of loci in strong linkage disequilibrium

Chromosome	Loci (D' score)		
1	47-48 **(0.89)**		
2	-		
3	-		
4	-		
5	2-3 **(0.85)**	39-40 **(0.96)**	54-55 **(0.98)**
6	22-23 **(0.77)**	66-67 **(1.0)**	
7	54-55 **(1.0)**		
8	49-50 **(1.0)**		
9	-		
10	43-44 **(1.0)**	46-47 **(1.0)**	
11	-		
12	1-2 **(0.79)**	9-10 **(0.98)**	
13	-		
14	23-24 **(1.0)**	48-49 **(0.79)**	
15	51-52 **(0.77)**		
16	-		
17	-		
18	-		
19	5-6 **(0.91)**		

Table 2.3. Loci pairs that have a D' score above 0.75. The numbers with hyphens represent the numerical positions of loci which the LD occurs between. Numbers in brackets are the D' score for each locus pair.

Hardy-Weinberg Equilibrium and Population Differentiation

Allele Differences

The number of monomorphic loci, and the mean within-site number of nucleotide differences varied greatly among sample sites, (Table 2.4). For example, mice from the SP site had the most monomorphic loci, with 902 (77%), and there were a similarly high number of monomorphic loci (857 loci, 74%) in mice from the SK site. The mice from the SK site also had the lowest mean number (140.6) of within-site nucleotide differences. Mice from the BM site had the fewest monomorphic loci at only 217 (19%); the greatest mean number (523) of nucleotide differences within a site was found in mice from the ST site. This suggests that the mice from SK and SP have relatively little genetic diversity compared to mice from ST or BM which may be due to a small population size, genetic drift, and inbreeding occurring at these sites.

HWE

Hardy-Weinberg equilibrium tests showed that when treated as a single population the allele frequencies of 1132 (88%) of all 1183 loci significantly deviate from HWE expectations (<0.05). To address the possibility of the Wahlund effect (Wahlund, 1928) occurring in these data, these analyses were then performed for mice from each sample site separately. On average only 5 % of loci deviated from HWE expectations when calculated per site. Mice from WF had the highest percentage of loci (12%) whose allele frequencies significantly diverged from HWE expectations. These HWE analyses provide evidence for the Wahlund effect and therefore of population sub-division and so from here-on sample sites will be referred to as sub-populations.

	No. Mice	Total No. Polymorphic Loci	% of Loci out of HWE	No. Monomorphic Loci	% Monomorphic Loci	% Heterozygosity	Mean Within-site No. nucleotide Differences
All Sites	433	1168	88.	136	11.6	-	-
BM	33	951	10.	217	18.6	21.1	437
GL	30	749	1.	419	35.9	22.6	394.6
HW	167	707	6	461	39.5	16.2	296.5
JB	36	784	11	384	32.9	15.3	305.6
LU	18	901	4	267	22.9	26.5	502.2
PF	10	491	3	677	58	17.7	285.4
PH	63	559	6	609	52.1	16.4	291.1
SK	30	311	3.	857	73.4	9.5	140.6
SP	6	266	0	902	77.2	12.1	165.7
ST	15	910	5	258	22.1	26.4	523.1
WF	18	733	12	435	37.2	15	349.4
WT	7	539	0	629	53.9	17	298
Mean		658.4	5.2	509.6	43.6	18	332.4

Table 2.4. A summary of several population genetic analyses from the multi-locus genotype data for each of the mouse sample sites and the mean of each measure across all mice and loci.

Pairwise F_{ST} and Nucleotide Differences

Pairwise F_{ST} values among sample sites were generally high (range 0.222-0.750; mean 0.461), (Tables 2.5. A-D), suggesting a high degree of genetic differentiation among mice from all sample sites. All pairwise F_{ST} comparisons were highly significant overall $p<0.0001$, suggesting that each sample site was a genetically distinct sub-population. However, individual loci differed in their F_{ST} values and their significance values. The repeated measures F_{ST} analyses also showed that all pairwise F_{ST} analyses were significant, $p<0.005$, and the general pattern of genetic differentiation remained the same, (Table 2.5). The similarity of the repeated measures results and the initial analyses show that despite differences in sample sizes among the sites the mice from these sites are strongly genetically differentiated.

The highest F_{ST} values were in all comparisons that included mice from site SK, with all of these having an F_{ST} greater than 0.5, and an average of 0.56. The highest F_{ST} value was 0.75, between mice from SK and SP. All comparisons involving mice from site ST consistently had the lowest F_{ST} values, with an average of 0.34, with the lowest overall value of 0.22 between mice from ST and LU.

The average number of pairwise nucleotide differences among sample sites shows a similar trend to the F_{ST} values, (Tables 2.4. & 2.5). The average number of nucleotide differences among mice from different sites was 667.9, ranging from 603.7 between BM and ST, to 778.6 between SK and JB.

The average number of nucleotide differences among wild mice and laboratory mice (LB) was 988.5, which were much higher than among wild mice, which was 332.4. Comparing wild and laboratory mice, the greatest difference (1004.7) was between laboratory mice and mice from SK, and the lowest (974.1 between laboratory mice and mice from WT.

A)	BM	GL	HW	JB	LU	PF	PH	SP	SK	ST	WF	WT
BM	0											
GL	0.347	0										
HW	0.465	0.447	0									
JB	0.461	0.434	0.525	0								
LU	0.307	0.271	0.452	0.42	0							
PF	0.417	0.402	0.493	0.525	0.354	0						
PH	0.458	0.422	0.503	0.508	0.415	0.522	0					
SP	0.446	0.437	0.545	0.564	0.383	0.574	0.553	0				
SK	0.564	0.568	0.603	0.661	0.564	0.68	0.63	0.75	0			
ST	0.228	0.265	0.377	0.401	0.222	0.323	0.408	0.371	0.541	0		
WF	0.402	0.365	0.474	0.498	0.349	0.476	0.476	0.538	0.653	0.33	0	
WT	0.384	0.364	0.491	0.509	0.325	0.477	0.503	0.573	0.684	0.3	0.45	0

B)	BM	GL	HW	JB	LU	PF	PH	SP	SK	ST	WF	WT
BM	0											
GL	0.216	0										
HW	0.331	0.379	0									
JB	0.317	0.336	0.464	0								
LU	0.224	0.238	0.402	0.349	0							
PF	0.348	0.385	0.483	0.465	0.381	0						
PH	0.348	0.371	0.500	0.453	0.387	0.518	0					
SP	0.421	0.467	0.590	0.546	0.451	0.591	0.600	0				
SK	0.477	0.540	0.637	0.610	0.528	0.641	0.650	0.742	0			
ST	0.110	0.204	0.288	0.301	0.219	0.323	0.344	0.413	0.478	0		
WF	0.372	0.397	0.517	0.492	0.413	0.534	0.528	0.635	0.687	0.361	0	
WT	0.323	0.350	0.484	0.451	0.360	0.478	0.501	0.579	0.635	0.314	0.51	0

C)	BM	GL	HW	JB	LU	PF	PH	SP	SK	ST	WF	WT

BM	0											
GL	0.319	0										
HW	0.394	0.44	0									
JB	0.428	0.46	0.542	0								
LU	0.250	0.27	0.390	0.402	0							
PF	0.391	0.43	0.488	0.535	0.361	0						
PH	0.407	0.43	0.519	0.528	0.373	0.523	0					
SP	0.474	0.51	0.604	0.625	0.436	0.585	0.607	0				
SK	0.531	0.58	0.647	0.680	0.527	0.640	0.654	0.745	0			
ST	0.227	0.29	0.337	0.408	0.233	0.342	0.388	0.444	0.515	0		
WF	0.331	0.34	0.429	0.468	0.296	0.436	0.431	0.531	0.595	0.310	0	
WT	0.377	0.41	0.512	0.532	0.354	0.485	0.518	0.589	0.650	0.347	0.416	0

D)	**BM**	**GL**	**HW**	**JB**	**LU**	**PF**	**PH**	**SP**	**SK**	**ST**	**WF**	**WT**
BM	0											
GL	0.255	0										
HW	0.357	0.401	0									
JB	0.333	0.365	0.463	0								
LU	0.266	0.290	0.428	0.387	0							
PF	0.360	0.402	0.476	0.464	0.409	0						
PH	0.361	0.392	0.492	0.449	0.409	0.507	0					
SP	0.446	0.483	0.595	0.552	0.485	0.582	0.595	0				
SK	0.498	0.557	0.644	0.621	0.565	0.642	0.649	0.75	0			
ST	0.145	0.241	0.311	0.318	0.263	0.323	0.352	0.43	0.494	0		
WF	0.323	0.343	0.458	0.430	0.377	0.470	0.451	0.56	0.630	0.320	0	
WT	0.327	0.363	0.477	0.441	0.386	0.462	0.486	0.57	0.633	0.310	0.43	0

Table 2.5. Pairwise FST values for (**A)** all mice and all loci, and **B-D** the repeated measures analyses 1-3 respectively. In A all values are highly significant $p<0.0001$, and in B-D all values are significant $p<0.05$.

AMOVA

AMOVA analysis showed that 48.2% of genetic variation is due to differences among sub-populations, 51.8% due to differences among individuals in the total population, and only 0.027% due to differences among individuals within a sub-population. For the three-group analysis, among Bristol/Gloucester, SK and LU mice, total F_{ST} rose to 0.52, with 41.5% of the variation due to differences among the sub-populations within a group, 47.6% due to variation among individuals within a sub-population, and only 10.8% of variation due to the difference among the three groups of sub-populations. The four-group analysis had an F_{ST} value of 0.49, with 44.5% of variation due to differences among sub-populations within a group, 50.5% due to differences among individuals within a sub-population, and 5.0% due to differences among groups of sub-populations. Together these results show that very little of the variation among wild house mice derives from the differences among groups of sub- populations. In contrast a very large proportion of the genetic variance among these mice can be attributed to the differences among the sub-populations themselves. These findings support the island model (Latter, 1973) of genetic sub-division where sub-populations of individuals exchange migrants equally to all other sub-populations and receive the same amount back, therefore the sub-populations become equally genetically sub-divided from one another. If there were larger contributions to genetic variance from groups of sub-populations this may suggest that those sub-populations within a group exchange migrant more frequently than to sub-populations out of the group.

Population Structuring

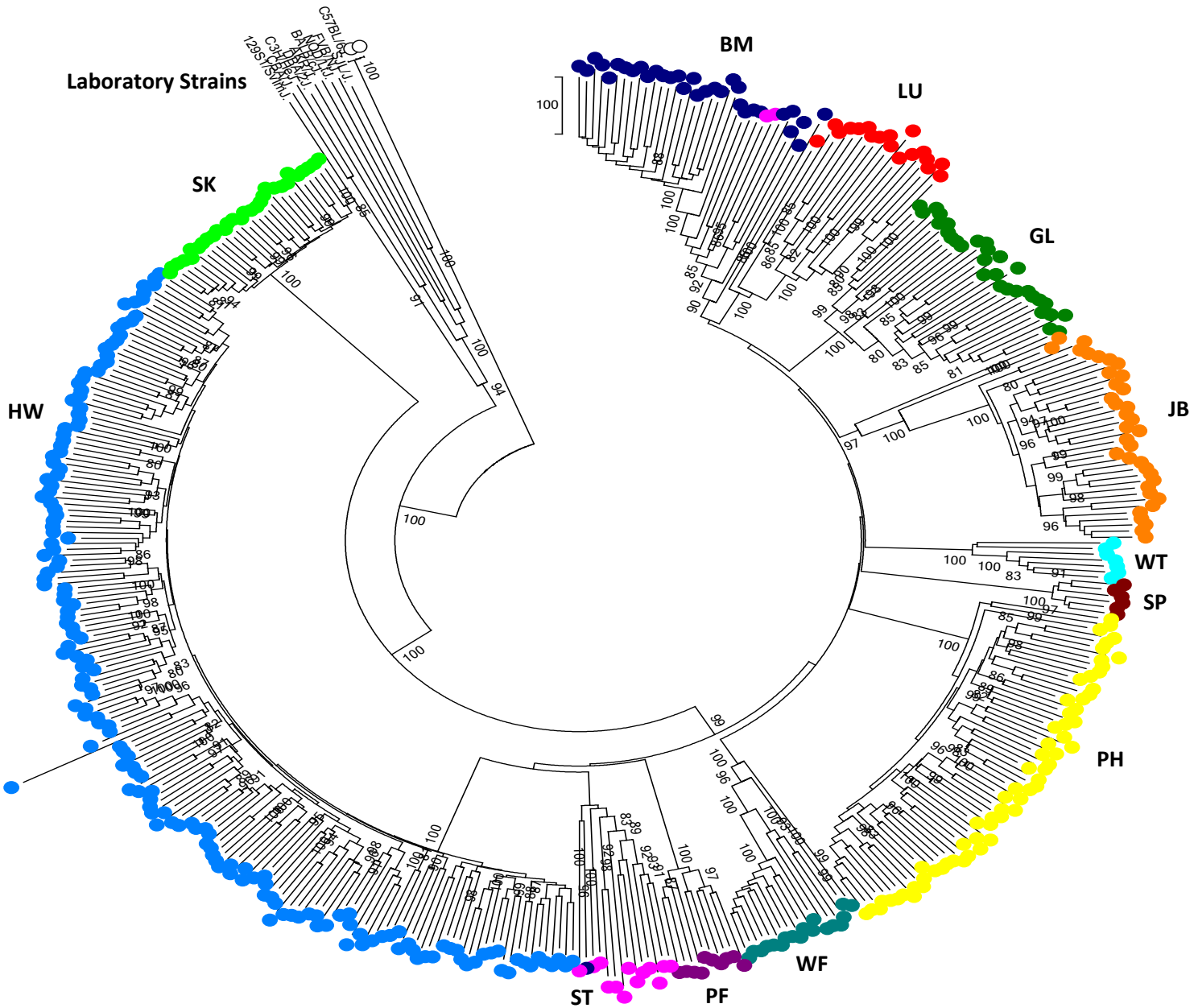
The NNJ tree (Figure 2.2) showed that the principal pattern was that mice from one sample site were generally close together in the tree, usually deriving from a single common branch of the tree, henceforth referred to as a "site-branch". These site-branches were highly supported as shown by the bootstrap values being greater than 80%. The relative positioning if site-branches to one another was not well supported. In general, the lengths of the site-branches are similar suggesting approximately equal separation among them. However, the site-branch of the mice from SK was considerably longer suggesting that these mice are the most differentiated form all other mice.

The NNJ tree shows that laboratory house mouse strains are branched together, but with very long branches connecting them to the wild mice. The average number of nucleotide differences among laboratory mouse strains was 1039.2. The laboratory control mice clustered with C57BL/6 mice on the same branch. The mean nucleotide distance between these two mice and C57BL/6 was 1. This further supports the accuracy of the genotyping of these mice. These branches were highly supported by the bootstrapping values. This analysis highlights the strong genetic differentiation that exists between wild and laboratory strains of house mice.

The positioning of certain wild mice in this tree was consistent with them being migrants between sample sites, specifically between the BM and ST sites which are <200m apart. For, example mouse 149 was sampled at BM, but clustered among mice from ST. Two mice, 78 and 145, were sampled from ST but clustered among the mice from BM. These three mice are therefore likely to have migrated between these two sample sites.

Mouse 306 was sampled from HW and clusters among other HW-sampled mice. However, the branch length of this mouse is much longer than other mice from the same site. The average nucleotide difference of this mouse to all other mice from HW was 675.0, compared to an average of 107.1 for all mice from this site. It may be that mouse 306 is a migrant from an adjacent sub-population that has not been sampled, but which is genetically more related to mice from HW than mice from any other site.

Among mice from the same site, those from SK had the shortest branch lengths, while BM mice had the longest. BM mice also seem to have two main sub-branches within the site-branch. It is notable that mice sampled from LU cluster together, but, among the site-branches of the Bristol/Gloucester mice. Mice from SK cluster the most distinctly from all other mice. These results further suggest that the wild house mice are strongly genetically sub-divided but that the relative amount of sub-division among most sub-populations is approximately equal.



Site Code	Colour
BM	
GL	
HW	
JB	
LU	
PF	
PH	
SK	
SP	
ST	
WF	
WT	

Figure 2.2 A nearest-neighbour joining tree of sampled and genotyped wild and inbred laboratory house mice and reference genotypes of inbred strains of house mice. The scale of the tree shown represents 100 nucleotide differences among individuals. Bootstrapped values are only shown for branches that have >80%bootstrap support.

STRUCTURE

The initial STUCTURE analyses of the wild mice showed that they have a strongly divided population structure. This was shown by mice sampled from one site deriving the majority of their genotype from a single common cluster. There was also a small amount of admixture present in many of the mice. However, three of the mice sampled showed higher levels of admixture compared the other mice, and the predominant cluster of these mice differed from the other mice sampled from the same site. This evidence suggests that these mice are migrants between the BM and ST sample sites; this will be discussed further, below.

From the initial STRUCTURE analysis the most likely value of K was determined to be K=4, as determined by the Evanno method (Evanno *et al.*, 2005) using STRUCTURE harvester (Earl and Bridgett, 2012). However, as the sample sizes from each sample site were uneven, they likely biased the reliability of STRUCTURE runs towards the clusters for the sub-populations with the highest number of samples. To compensate for this a repeated measures STRUCTURE analysis using only six samples from each site was carried out, as described above. These analyses were then compiled, and the most likely value of K was again determined using STRUCTURE harvester. The STRUCTURE harvester analysis of these data suggest that K=9 is the most likely value of K. This result better reflects the pattern of clustering of individuals that occurs as values of K increase. The pattern of clustering itself reveals that mice from SK are the first to cluster separately and therefore are the most genetically distinct from the other mice.

Examining the main cluster plots from CLUMPAK, in general all individuals sampled from the same site shared members of the same cluster in both types of STRUCTURE analysis. There was generally a low level of admixture and of migration. The initial STRUCTURE analysis resolved largely in order of sample size as K values increased. The first sub-population to be resolved was HW, the sub-population with the largest sample size, at K=2. Some admixture from this cluster was also observed in all other sub-populations, excluding PH and SK. The next sub-population to resolve was PH at K=3, the second largest sample size, followed by SK at K=4, the sub-population with

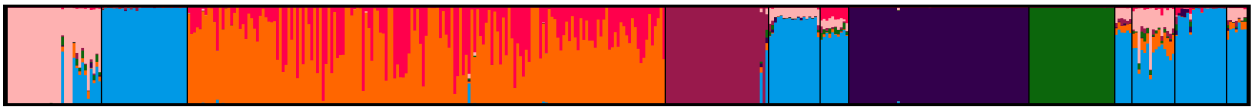
the highest pairwise F_{ST} values to others. At K=5 and K=6 the JB and BM sub-populations resolve respectively. From K=7 onwards the HW sub-population is an admixture of the K2 cluster and the K7 cluster and briefly experiences admixture of the K10 cluster at K=10. Sites continue to resolve primarily in sample size order as K values increase to K=12, see figure 2.3. The WT and ST sub-populations were the last to resolve with mice from WT the only sub-population not to have membership to a single cluster. The minor cluster plots of each K value are similar in pattern to the major cluster plots with small differences in the admixture among sub-populations, see figure 2.3.

As mentioned above some admixture was observed within sub-populations. The most prominent admixture is seen in the HW mice. At K=12 other notable amounts of admixture can be seen in the BM, GL, SP, ST, WF and WT. In the BM and ST sub-populations there are three mice that derive over 50% of their genotype from the other clusters suggesting these mice may be migrants, as previously shown by the NNJ tree.

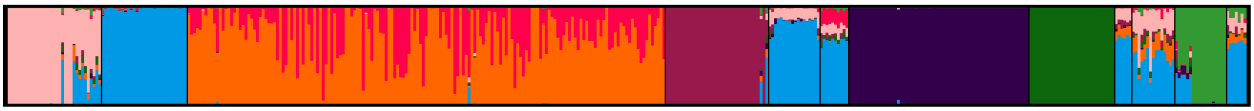
In comparison to the initial STRUCTURE analysis, CLUMPAK plots of the repeated measures STRUCTURE analyses did not resolve in order of sample size but followed a trend that was closer to the pairwise F_{ST} values among sub-populations, see 2.2. The first sub-population to resolve in the repeated measures analyses was SK at K=2, the sub-population with the highest pairwise F_{ST} values compared to other sites, this was followed by SP at K=3. As K values increase more sub-populations are resolved however with relatively high amounts of admixture and less distinction compared to the initial analysis until K=11 where, apart from BM and ST, individuals from each sub-population belong to a unique cluster. The BM and ST sub-populations are last to resolve, which mirrors the relatively low pairwise F_{ST} values among these sub-populations and the smallest geographical distance between them. The admixture observed prior to K=11 is a result of STRUCTURE trying to assign membership to individuals when their unique cluster is not available.

BM	GL	HW	JB	LU	PF	PH	SK	SP	ST	WF	WT

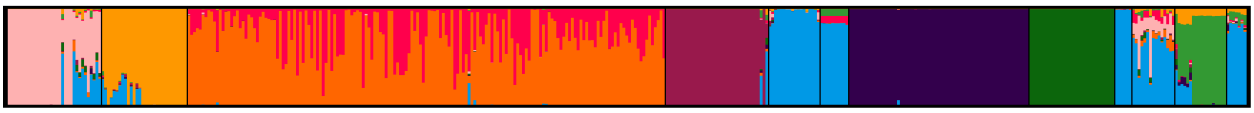
K=7



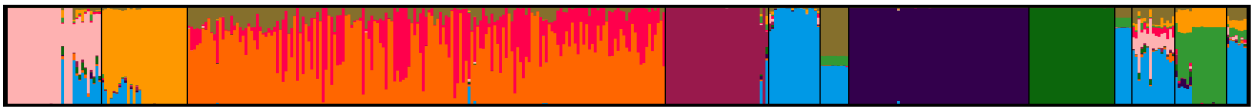
K=8



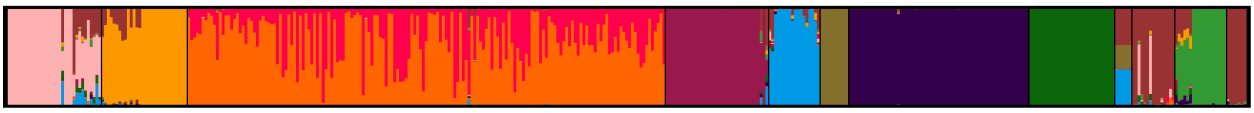
K=9



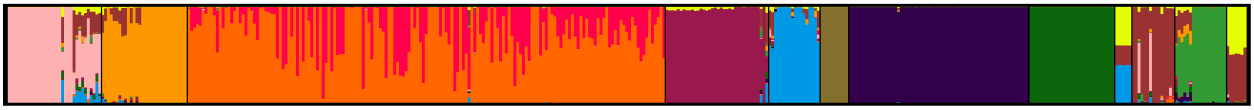
K=10



K=11

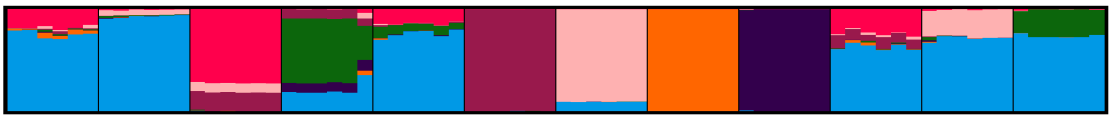


K=12

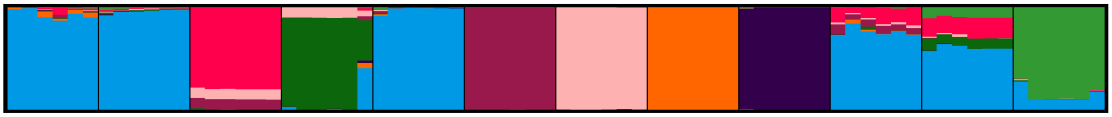


BM	GL	HW	JB	LU	PF	PH	SK	SP	ST	WF	WT

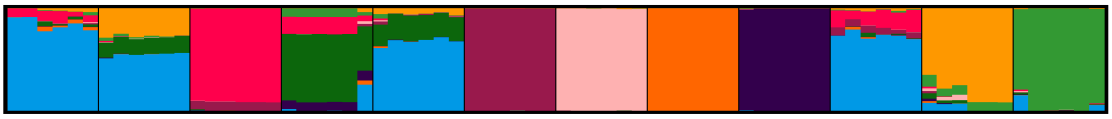
K=7



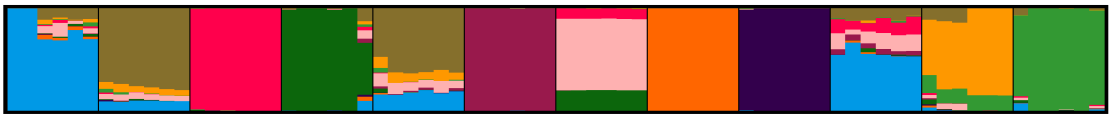
K=8



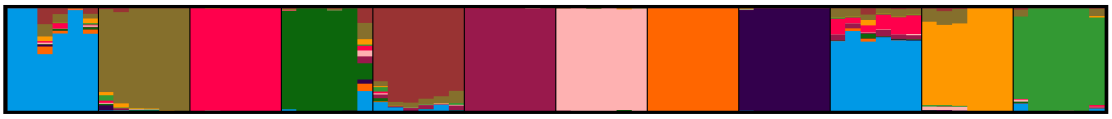
K=9



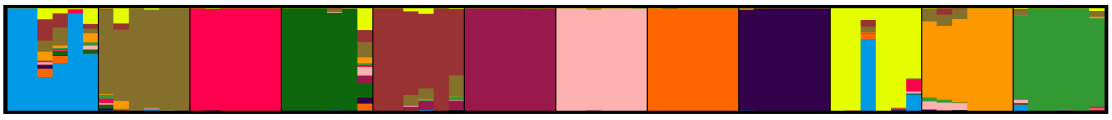
K=10



K=11



K=12



C

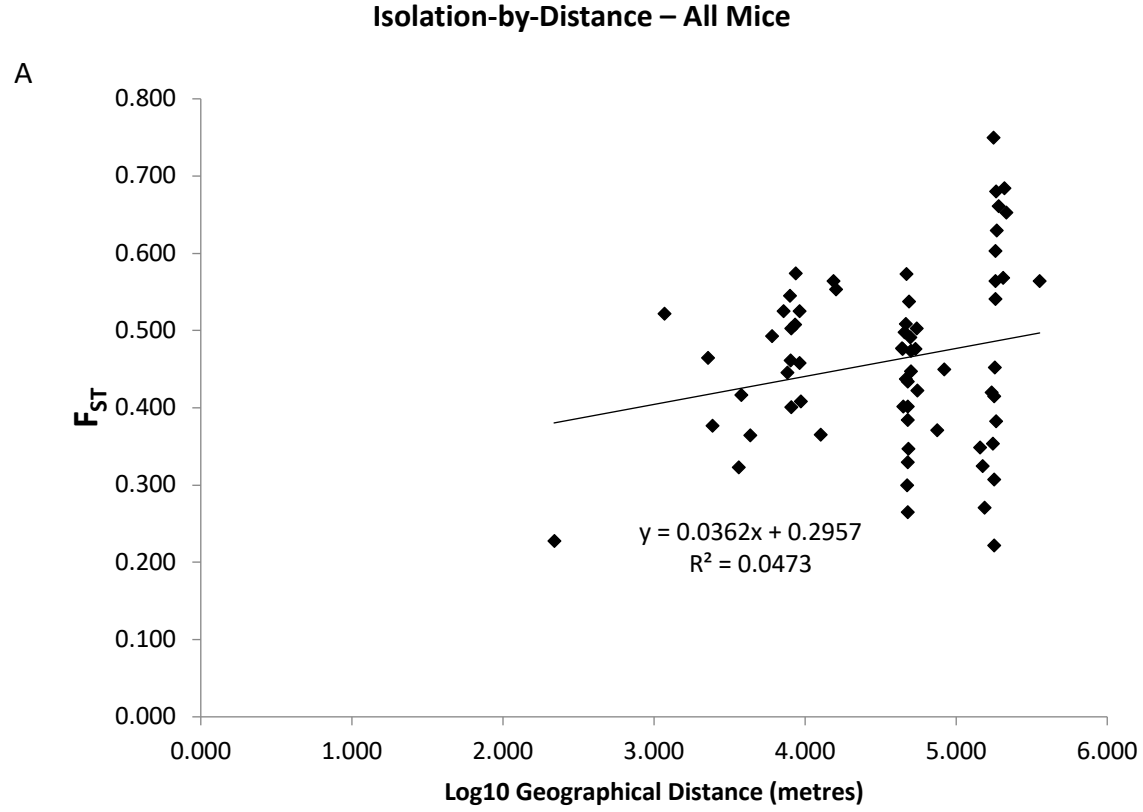
	Initial	Repeated
K=7	10/15	11/15
K=8	6/15	10/15
K=9	14/15	12/15
K=10	10/15	10/15
K=11	8/15	7/15
K=12	10/15	5/15

Figure 2.3. The results of the (A) initial STRUCTURE analysis and (B) repeated measures STRUCTURE analysis compiled and plotted in CLUMPAK. Both figures (A and B) the major cluster plots for K values of 7-12. In total there were 15 repeats in each analysis and (C) shows the number of repeats that each displayed plot is representative of. The mice are ordered by the site they were sampled from.

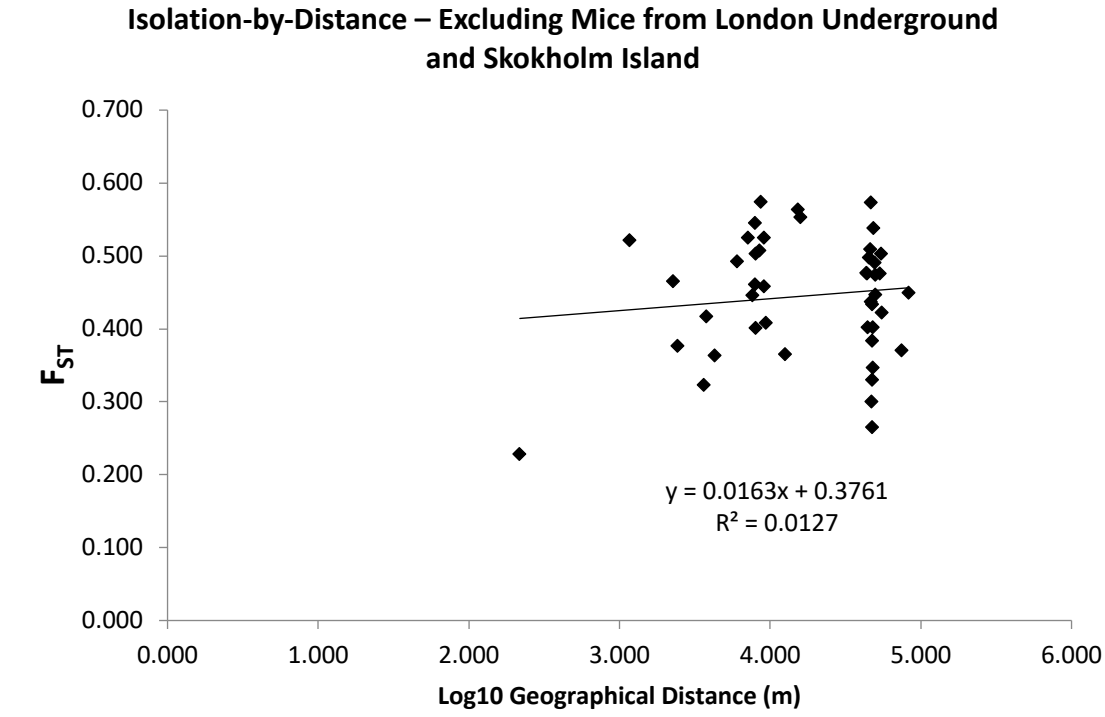
Isolation by Distance

No significant isolation-by-distance was found between the geographic distance (log10 meters) and genetic distance (F_{ST}) when all mice were included in the analysis (R*xy* = 0.217, *p*= 0.21) (Figure 2.4.A). There was also no significant correlation when mice from London underground and Skokholm island were removed from the analysis (R*xy* = 0.113, *p*= 0.39) (Figure 2.4.B). However, significant isolation-by-distance was observed when only mice from the Bristol sample sites (BM, HW, JB, PF, PH, SP and ST) were included in the analysis (R*xy* = 0.448, *p*= 0.01)

A


Isolation-by-Distance – All Mice
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
F_{ST}
y = 0.0362x + 0.2957
R² = 0.0473
0.000
1.000
2.000
3.000
4.000
5.000
6.000
Log10 Geographical Distance (metres)

B


Isolation-by-Distance – Excluding Mice from London Underground and Skokholm Island
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
F_{ST}
y = 0.0163x + 0.3761
R² = 0.0127
0.000
1.000
2.000
3.000
4.000
5.000
6.000
Log10 Geographical Distance (m)

C

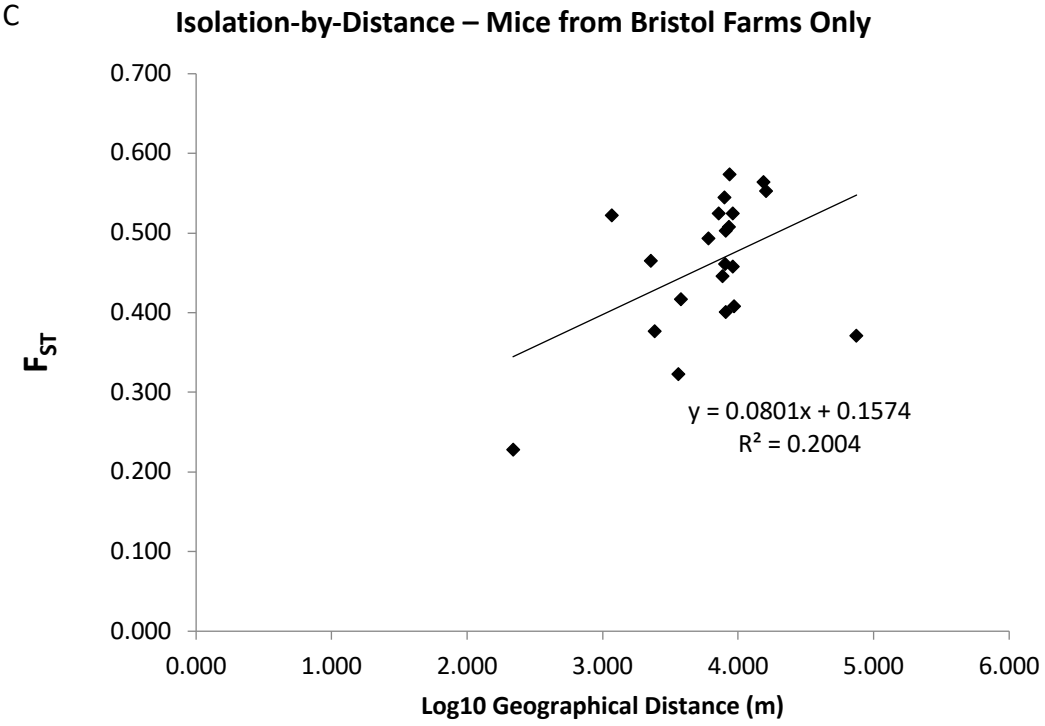


Figure 2.4. Mantel test for isolation-by-distance for all mic (A), All mice excluding from LU and SK (B) and for Bristol only mice (C)

Discussion

Population genetic studies of wild house mice have been fairly scarce over the past 50 years. In contrast, laboratory inbred-strains of house mice have been the bedrock of much of modern biological research. The tools allowing detailed population genetic analyses of wild house mice have been developed as a by-product of these studies of laboratory mice. However most recent studies of wild house mice have primarily focused on populations across large geographic distances, between countries or continents, and the phylogeny of sub-species of wild house mice, rather than the population genetics and population structure within a single population (Bonhomme *et al.*, 2007). In the current study I have utilised a house mouse-specific genotyping tool, developed for comparisons of laboratory inbred-strains of mice, to generate genotypic data for wild house mice sampled from Southern Britain. Analyses of these data reveal that these live in highly structured sub-populations. The amount of genetic differentiation among these sub-populations is far higher than that estimated in previous studies of wild house mice, and is comparable to differences among domesticated breeds of animals including among house mouse sub-species (Lyimo *et al.*, 2014; Murakami *et al.*, 2014; Uzzaman *et al.*, 2014; Alves *et al.*, 2015; Iacolina *et al.*, 2016).

Neutrality and Linkage Disequilibrium

Both non-neutral loci and loci exhibiting high linkage disequilibrium may bias other population genetic analyses (Helyar *et al.,* 2011). The vast majority of loci utilised in this study were found to be neutral with only 15 deviating from neutrality. This finding is not surprising as the SNP loci genotyped in the Mouse MD Linkage assay were selected to distinguish among laboratory strains of house mice and not for functional genetic variants, therefore they were assumed to be neutral.

Linkage disequilibrium was shown to be low among the loci examined in the current study. The average physical distance between adjacent genotyped loci was ~300kb. Laurie *et al.* (2007) suggest that in wild house mice the physical distance between markers required for strong linkage disequilibrium to occur is less than 100kb (Laurie

et al., 2007). Therefore, the distances among loci in the current study far exceed this and LD would not be expected to occur. The absence of strong LD among loci and the removal of loci that are likely under selection ensures that the analyses carried out on the remaining genotypic data should not be biased.

Genetic Differentiation and Population Structure

All population genetic analyses (HWE/Wahlund effect, F_{ST}, AMOVA, and STRUCTURE) show that the sampled wild house mice live in highly structured, genetically differentiated populations, to the extent that they should be considered distinct sub-populations. Sub-divided populations become more genetically diverged from one another over time, both because of the genetic drift of each sub-population and because of a lack of gene flow among them. HWE results indicate that genetic sub-division exists, and to determine how much division exists, and among which mice, F_{ST} and AMOVA analyses were conducted. For these analyses each sample site was considered a distinct sub-population.

AMOVA results indicate that almost 50% of the genetic variation among the wild house mice derives from differences among this sub-populations. This population genetic structure of wild mice means that a high amount of within-sub-population inbreeding occurs. This has the effect of reducing heterozygosity of individuals within-sub-populations without reducing the genetic diversity of the population overall. The observed heterozygosity of these sub-populations is therefore lower than is expected from the genetic diversity, heterozygosity, of the total population, giving a high F_{ST} values for pairwise comparisons.

Comparing mice from different sample sites by F_{ST} and by the number of nucleotide differences, all show that that mice from the SK sub-population are the most inbred and genetically isolated, compared with from mice from all other sample sites. This result is fully consistent with the ecology and biology of this population. The SK sub-population was founded over 100 years ago and migration to and from the island is negligible (Berry, 1964; Berry and Jakobson, 1974). This will result in the genetic isolation and inbreeding. As a feral population of house mice, there are also less

restrictions to migration across the island with a less stringent territoriality and breeding system, which homogenises the allele frequencies across the sub-population. STRUCTURE analyses show that mice from the SK sub-population have the least admixture of all the sub-populations, differentiating as a distinct cluster at K=4 in the initial analysis. In further support of this is the NNJ tree, where the mice from SK have the shortest within-site branches, the site-branch itself is the longest, among the wild sub-populations, highly supported by bootstrapping values.

In contrast to mice from SK, wild mice from LU were some of the most heterozygous and genetically diverse. Although almost geographically equi-distant from the Bristol/Gloucester sub-populations as the SK sub-population, mice from LU are genetically much more similar to the Bristol/Gloucester mice than the mice from SK are. Initially it was thought that the location of these mice, in tunnels and platforms underground, would genetically isolate them and prevent migration and therefore gene flow to other sub-populations. However, the relatively high levels of heterozygosity, and the high number of nucleotide differences among the mice within the sub-population suggests that migration among sub-populations occurs more frequently than in other sub-populations. These mice live on a mass transport network of the most well-connected city in the country. It is possible that this assists the migration of mice to and from the LU sub-population. The mice in this study were sampled from three different stations on the London Underground and from only one platform at each. As the number of individuals sampled was low they were considered a single sample site (now sub-population) and analysed this way. STRUCTURE and NNJ tree analyses did not suggest further sub-division of this sub-population. A larger scale genetic study of mice from LU could give greater insight into the migration and gene flow of these mice. In particular it may be of interest to genotype mice from different platforms at the same station or different stations along the same train line and examine the gene flow among these individuals.

The HW sub-population had the largest sample size of wild mice from a single sample site. The mice cluster strongly on the HW site-branch of the NNJ trees and the mice from HW have the third lowest average number of within-site nucleotide differences.

However, STRUCTURE analysis above K=6 reveal that the HW mice are admixed. There are two distinct genotype clusters that occur and they are exclusive to the HW mice, and the amount of admixture in these mice varies among individuals. Surprisingly, the genotype clusters that are present in the mice from HW are not also shared with mice from BM or ST as these mice are less than 2.5km away. This may indicate that this distance exceeds the limits of dispersal and migration of house mice. Significant isolation-by-distance was not observed among the sub-populations of the current study. It is likely that beyond a certain distance no migration occurs and therefore the correlation between genetic similarity and geographic distance ceases. From the current study it is impossible to determine whether the genotypic clusters observed among the mice from HW are a result of interbreeding of two distinct sub-populations not sampled in this study or whether the mice sampled are gradually diverging from one another. In the current study the specific locations of mice sampled within the HW site were not recorded. This data may assist in identifying the cause of the genetic admixture among the sampled mice as it may show a gradient of gene flow among the mice or identify new sub-populations all together.

Migrants and Admixture

The NNJ tree identified the presence of three possible migrant mice, see figure 2.2 However, STRUCTURE analyses reveal that these three individuals have genomes that are mixtures of the two relevant populations. This suggest that these individuals may not be first generation migrants, but that their parents / grand parents may have migrated. In this case, among the 433 mice, we observed no contemporaneous migratory mice. Admixture is also observed, albeit at a much lower extent, in other mice from the BM site. This suggests that migration and gene flow between these two sub-populations can occur but is still rare and was not directly observable when the mice were sampled. If we consider the three possible migrant mice as first-generation migrants, they would account for only 6.25% of the wild house mice sampled from these two sub-populations. This alone may not be sufficient gene flow for allele frequencies in these two sub-populations to converge, but there are also two further challenges for migrant mice to overcome. Firstly, the breeding behaviour and territorial structure of resident house mice may restrict the integration of migrant

mice into the resident sub-population. Migrant commensal house mice are often attacked and killed by resident mice (Crowcroft and Rowe, 1963). Thus, the genetic analyses will only assay the effective migration rate, not the actual migration rate. Secondly, the instability of these sub-populations may inhibit the convergence of allele frequencies between them. A study of commensal house mice by Baker (1980) suggested that commensal house mouse populations are only stable for ~18 months. Baker observed that sub-populations were established by a few founding members, reproduce for up to nine generations, after which the populations would crash and the cycle start again (Baker, 1981). This results in repeated founder effect events. These founder effects can have a significant effect on the allele frequencies of a population, particularly populations that are short lived (Berry, 1967). Overall, this has the effect of preventing allele frequencies converging among sub-populations.

Causes of Population Sub-division

The amount of genetic variation that has been observed among the wild house mice in this study is equivalent to that observed among breeds of domesticated animals, which have been in reproductive isolation for hundreds if not thousands of years. The large F_{ST} values observed in the current study are much larger than those reported in previous population genetic studies of wild house mice (Petras, 1967a; Searle *et al.*, 2009; Gabriel *et al.*, 2013). This may be due to differences in the loci and the populations being investigated; previous studies primarily used a handful of loci and populations in very close proximity to one another (Anderson, 1964; Petras, 1967a). However, changes in human activity may also be a cause of the genetic isolation among sub-populations. Over the past 50 years there has been a substantial decrease in the amount of hedgerows, ditches and dykes that exist on British farmland (Kotzageorgis and Mason, 1997). Previous studies have suggested that these fringe habitats provide an important population reservoir for small mammals, including house mice (Rowe *et al.,* 1987). Mice from commensal habitats periodically migrate and breed in these fringe sub-populations. Following population crashes; animals migrating from the reservoir populations rapidly repopulate commensal sites (Rowe *et al.*, 1987). Hedgerows, dykes and ditches provide cover for migrating house mice and allow for gene flow to indirectly occur among sub-populations located further

apart than that which an individual mouse can feasibly migrate. The loss of these habitats may therefore have further genetically isolated the sub-populations of wild house mice, as we have observed here.

The geographical distances among sub-populations are a barrier to gene flow. It would therefore be sensible to expect that in this study there would be significant isolation-by-distance considering the location of the sites that were sampled. However, only when mice from the Bristol sites were considered alone was significant Isolation-by-distance observed. Despite the SK and LU sites being considerable geographic distances from the other sites their genetic distances were not proportionally much greater, resulting in no significant correlation. The high genetic distances among relatively close sites (geographically) together with the lack of significant IBD supports the theory of the deme breeding structure in wild mice. An individual mouse typically migrates less than 100m in a lifetime and, limiting the effective geneflow among sub-populations to short distances. Beyond a threshold distance, estimated to a few kilometres, geneflow is very low and effectively equal to all sites beyond the threshold. As the sampling of sites in the current study was sporadic it is difficult to estimate a more accurate distance of this geneflow barrier threshold. Future studies may wish to sample mice at smaller distance intervals in order to identify this distance. These findings also suggests that in general the wild mouse population conforms to an island model (Latter, 1973) with mice living in isolated "island" populations with rare, effective migration among them. It is likely that isolation-by-distance does occur among commensal house mice on a smaller scale, (i.e. within a single farm building), than has been examined in this study. There is likely to be a gradient of gene flow among mice inhabiting farm buildings with sub-populations, with more proximally located sub-populations exchanging reproductive members more frequently than those further apart. Beyond a certain distance the gradient of this gene flow becomes negligible. Re-sampling sites to give a greater number of mouse samples and recording the specific location of each sample within a site may reveal this.

The high reproductive rate of house mice is also a factor contributing to the significant population genetic sub-division observed. Rapid generation times of this species

means that multiple generations can occur in a single year, and this further exacerbates the rate of genetic drift that occurs between sub-populations. Adding to this is a high turnover rate of the sub-populations themselves. As previously discussed, commensal house mouse populations are generally only stable for short periods of time (Baker, 1981), which is largely due to human activities such pest control or the clearance of equipment from buildings. As such, sub-populations are repeatedly crashing (sudden rapid declines) and being repopulated by a small number of individuals, resulting in declining gene pools. Over time, this causes genetic sub-division among sub-populations. This also limits the extent to which migration can maintain allele frequencies among sub-populations as the migration rate must be maintained at a high level to overcome the population crashes and allow allele frequencies to converge.

Conclusion

Wild house mice live in differentiated sub-populations. Mice from Skokholm Island are the most inbred and strongly differentiated from all other sub-populations. Mice from London Underground are relatively outbred with a high proportion of heterozygous loci compared to the other sub-populations. Despite this, mice from LU are not as genetically distinct from mice from the Bristol/Gloucester sub-populations, despite their substantial physical separation from southwest England. Little genetic admixture exists among most of the sub-populations and migrants are rare even among neighbouring sub-populations. The amount of genetic differentiation among sub-populations is considerably higher compared to other mammal species and is comparable to the genetic differentiation of domesticated breeds of livestock. The large amount of genetic differentiation that exists among sub-populations of wild house mice is likely the result of house mouse migratory and reproductive behaviour, the instability of the commensal sub-populations, and possibly the result of human activities on farm landscape and agricultural practices.

However, there are a number of limitations to the population genetic analyses of the current study. Firstly, the sample sizes from each sample site were highly unbalanced with the smallest sample size being 7 (PF) and the highest 181 (HW). The sampling discrepancies were due to the unpredictable nature of live trapping a wild population. Sites were not sampled from at regular intervals and the number of mice trapped during each sampling attempt varied greatly. Therefore, any future study of multiple sampling sites should take this into account and an initial trapping phase be used to assess the reliability and population size at each sampling site. Secondly, the exact locations of the sampled mice from each site, specifically the distances among the trapping sites, was not recorded and so it was impossible to determine whether isolation-by-distance occurred on a small scale, such as within a single building or farm. A longitudinal study of a single sub-population may help in understanding the effect of population crashes on the allele frequencies and population genetics of wild commensal house mice over time. The HW sub-population lends itself well to this kind of study as it was the most reliable sampling site and experiences periodic population crashes due to pest control practices.

Chapter 3: The Heritability of Measures of Immune Function in Wild Mice

Abstract

The heritability of traits is often studied in controlled populations using a large number of genetic markers as well as large sample sizes. Currently there is not consensus on the heritability of immunological traits as there is so much variation between studies. Most studies of heritability have been carried out on humans or domestic livestock where relatedness is known and does not need to be estimated. Here a novel approach was taken to assess the heritability of 18 immunological traits in a population of wild house mice. The relatedness of these mice was not known and needed to be inferred from their genetic similarities using the program PLINK. Once relatedness was estimated the heritability of these immune traits could be calculated in GCTA. Initially eight immunological traits were identified to have significant heritabilities. However, due to the strong population subdivision a further analysis of the largest population (containing over 180 individuals) was carried out. With the site effect removed and weight (as a proxy for age) and sex fitted as covariates six traits were found to be significantly heritable, three of which were common to the initial analysis. These finding again demonstrate the variability in calculating heritability of immunological traits, but also that several of these traits are highly heritable and under strong genetic control. In particular antibodies, specifically IgG and IgE are shown to be very heritable.

Introduction

Animals use their immune systems to protect themselves from potentially harmful infections. Resistance to pathogenic infection is of key importance to the survival of organisms, and therefore we would expect this feature to be selected for, and little variation in immune function among individuals of a species. However, measures of immune function, from here referred to as immunological traits, do vary greatly among individuals (Abolins *et al.*, 2011). The variation in these immunological traits is often a consequence of exposures to different infections (Brodin and Davis, 2017) as different pathogens require different immunological responses. In addition, past infections may modify host immune responses (Brodin and Davis, 2017) and organisms are often are infected with multiple organisms (Lello *et al.*, 2004; Damania and Dittmer, 2014). In spite of these influences the variation in immunological traits may also be influenced by genetic variation among individuals as seen in the heritability these measures (*i.e.* the proportion of phenotypic variation explained by genetic variation). Studies in humans, domesticated animals and even some wild animal populations provide evidence that many measures of immunological traits, such as cell counts, antibody titres and serum cytokine concentrations are heritable (de Craen *et al.*, 2005; Pitala *et al.*, 2007; Visscher *et al.*, 2008; Brodin *et al.*, 2015). Studies comparing laboratory-inbred strains of house mice, in which the environmental conditions are identical among strains, show that measures of immune function significantly differ between the strains (Lee YH and Kasper LH, 2004; Valdar *et al.*, 2006; Packiam *et al.*, 2010; Sellers *et al.*, 2012), but not between individuals of the same strain. As the environmental conditions are identical among these mouse strains the sole variable is their genetic background. This suggests that the phenotypic variation of these immunological traits among strains is heritable, though this heritability has rarely been formally quantified for laboratory mice (Valdar *et al.*, 2006) and is yet to be studied in wild mice.

In general, studies examining the heritability of measures of immune function have been limited to humans with a few examples in domesticated species (sheep and pigs) and nesting birds (Smith *et al.*, 1999; Clapperton *et al.*, 2008; Kinnard and Westneat,

2009). The paucity of such studies of wild populations is due to limitations in sample availability, genetic resources and immunological research tools for these species. However as both genetics and immune function of house mice have been intensely studied in research laboratories for almost one hundred years, the necessary research resources and tools already exist for this species. Therefore, wild house mice present an almost unique opportunity to determine the heritability of immunological traits in a wild, unmanaged mammalian population.

Heritability of Measures of Immune Function in Mice

Inbred strains of house mice live in highly controlled and standardised environments. The genetic background of these mouse strains is the primary variable in laboratory experiments comparing inbred strains of mice. Therefore, these laboratory mice lend themselves to heritability studies. Inbred strains of house mice have been recorded to differ in most aspects of their immunology (Reviewed in Sellers *et al.*, 2012). Despite genotype and environment being recognised as important sources of variation on the phenotype of complex traits in the house mouse ,these factors are rarely quantified (Valdar *et al.*, 2006). More often, these studies use assumed genetic differences among inbred mouse strains to infer that a trait is heritable, and to then conduct studies to identify the underlying causal loci (Marques *et al.*, 2011). Because house mice are used as models for human disease and biology, most of these studies focus on indirect measures of immunological traits such as disease presence/absence and immunologically mediated phenotypes, such as inflammation (Marques *et al.*, 2011). Therefore, for rodents very little is actually known about how heritable specific immunological traits are, especially in wild populations where environmental conditions are highly variable.

One of the few studies that has directly measured the heritability of immunological traits in house mice examined a population of individuals descended from crosses between eight inbred mouse strains (A/J, AKR/J, BALBc/J, CBA/J, C3H/HeJ, C57BL/6J, DBA/2j and LP/J) (Valdar *et al.*, 2006). These mice were assessed for physiological and environmental covariates, including for immunological traits; percentage of total (%) B220+, %CD3+, %CD4+, %CD4+/CD3+, %CD8+, %CD8/CD3+, %NK cells. %CD8+ and the

ratio of the proportion of CD4+ and CD8+ cells were the two most heritable traits (out of 88 measured) with narrow-sense heritabilities of 0.89 and 0.8 respectively. Moderate heritability was observed for %B220+ (0.60), %CD3+ (0.51), Lymphocyte cell counts (0.48), %CD4+ (0.41) and total white blood cell counts (0.41). %NK cells was the only immunological trait to have a low heritability at only 0.22.

Aims

- The aim of this study is to determine which immunological traits are significantly heritable in a wild population of house mice

Methods

DNA was extracted and genotyped from 443 wild house mice at 1449 loci. The DNA from these samples was then genotyped and then filtered to remove poorly genotyped loci and individuals, and also to remove non-neutral loci; see chapter 2. This resulted in 433 individuals and 1168 loci to use in a population design narrow sense heritability analysis. Immunological measures were provided from data obtained by Dr Stephen Abolins, University of Bristol, from the same sample of mice; (see methods in Abolins *et al.*, 2017). The immunological measures include the scaled cell counts (the number of cells scaled by the mass of the mouse) for $CD4^+$ T cells ($CD4^+$), $CD8^+$ T cells ($CD8^+$), $CD19^+$ B cells (B cells), NKp46 natural killer cells (NK cells), Ly6G neutrophils (Neut), CD11c dendritic cells (DCs) and F4/80 macrophages (M□). The total serum concentration of IgG, IgE, and concentration of faecal IgA, as well as the concentrations of several cytokines produced following *ex vivo* stimulation of cells extracted from the spleens. These cells were stimulated with four molecules: CpG, CD3/CD28, peptidoglycan (PG) and lipopolysaccharide (LPS). Cytokines were produced from the stimulations with all four antigens, however, for the heritability analyses only data from a single antigen was selected for each cytokine. The stimulation selection was based on sample size in order to provide sufficient statistical power for the heritability analyses. CD3/CD28 stimulation data was used for IFN-γ and IL-4, PG stimulation for IL-1b and IL-12p70, and for CpG stimulation IL-6, IL-10, IL12p40 and IL-13, see table 3.1. for the sample size numbers of each immune phenotype.

	Phenotypes																	
Site	**B cells**	**CD4**	**CD8**	**DC**	**M**	**Neut**	**NK**	**IFN-γ**	**IL-1b**	**IL-4**	**IL-6**	**IL-10**	**IL-12p40**	**IL-12p70**	**IL-13**	**IgG**	**IgE**	**IgA**
BM	17	14	14	13	12	13	16	10	10	10	10	10	10	10	10	19	19	16
GL	29	27	27	28	28	28	19	22	22	22	22	22	22	22	22	30	30	30
HW	118	108	108	108	108	108	115	64	63	64	63	63	63	63	63	151	151	91
JB	32	31	31	30	31	30	30	27	26	27	26	26	26	26	26	35	35	25
LU	9	8	8	9	9	8	7	3	7	3	7	7	7	7	7	7	7	11
PF	10	7	7	8	8	8	10	5	5	5	5	5	5	5	5	10	10	9
PH	56	56	56	55	55	55	50	35	35	35	35	35	35	35	35	63	63	57
SK	28	29	29	28	27	27	29	14	14	14	14	14	14	14	14	30	30	17
SP	3	5	5	5	5	5	5	4	4	4	4	4	4	4	4	6	6	6
ST	13	11	11	12	12	12	12	7	7	7	7	7	7	7	7	15	15	12
WF	16	16	16	15	15	16	13	15	15	15	15	15	15	15	15	15	15	5
WT	6	5	5	6	6	6	6	4	4	4	4	4	4	4	4	7	7	6
Total	337	317	317	317	316	316	312	210	212	210	212	212	212	212	212	388	388	285

Table 3.1. The number of samples used for each immunological trait by, site and total.

Analysis

All analyses were carried out in collaboration with Dr Gibran Hemani, University of Bristol. PLINK 1.07 (Purcell *et al.*, 2007) was used to filter the genotypic data, and loci that had minor allele frequencies of less than 5% were removed from the data file and excluded from further analyses. A principal component analysis (PCA) was then carried out using the remaining genotypic data and the eigenvector values from this were then used as a covariate of micro-population structure in the heritability analysis. GCTA v1.25.3 (Yang *et al.*, 2011) was then used to construct multiple genetic relatedness matrices (mgrm) of all mice using an identity-by-state method (Yang *et al.*, 2010), which is a proxy for identity-by-descent.

Each of the immunological phenotypes, listed above, was rank-transformed to normalise the data and this was then used with the mgrm data to carry out a heritability analysis using GCTA (Yang *et al.*, 2011) for each immunological phenotype. Due to the multiple sample sites studied (see chapter 2 figure 2.1) the common-site phenotypic variance (which can include both genetic and environmental variance) was first calculated and then the heritability of the trait was calculated from the remaining phenotypic variance. Several covariates were also included in the heritability analysis; specifically, sex, site of capture, and 10 eigenvector values from the PCA analysis which was used to capture any micro-population structure among the mice. A likelihood ratio test in GCTA (Yang *et al.*, 2011) was used to calculate the probability of heritability estimates being different from zero.

The heritability analyses were unable to converge for three immunological traits: the number of DCs, Neuts, and the faecal concentration of IgA. This was likely due to the small sample sizes (see table 3.1) and the degree of variation for these traits. The results of these analyses for each immunological phenotype give a (i) heritability estimate with a standard error (s.e), (ii) a *P* value which is the probability that the trait had a heritability greater than zero, and (iii) an estimate of the proportion (and s.e) of the phenotypic variance attributable to site effects. In order to calculate heritability

without the site effects more accurately an analysis of only mice from the Hyatt's wood sample site were used in a further analysis. The eigenvectors were removed from this analysis as there were no longer site effects in action, and weight was added as a covariate for age. Previous analyses show that age and weight are highly correlated in house mice (Abolins *et al.* 2017) and the data set for weight was more complete.

Results

The average site effect value, the proportion of variation attributable to variation in the sampling sites, was estimated to be 0.24. The immune phenotype most affected by site effects was IL-1b at 0.67 (s.e. 0.18) with the least affected $CD4^+$ cells, NK cells, IL-4 and IgG at ~$1x10^{-6}$ (average s.e. 0.39), (Table 3.1). The average heritability of the fifteen immune traits, calculated from the remaining phenotypic variance after site effects were removed, was 0.27. In total eight immunological traits had heritability values significantly greater than zero: B cells, NK cells, IL-1b, IL-4, IL-12p70, IL-13, IgG and IgE ($p<0.05$) The highest estimate of heritability was for the cytokine IL-4 at 0.50 (s.e. 0.31) and the lowest was for IL10 at 0.07 (s.e. 0.09), (Table 3.1)

However, despite 8 out of 15 immune traits having heritabilities greater than zero the degree to which these traits are heritable was generally low, half of these traits had heritability values of less than 0.25. Heritability values above 0.25 may be considered moderately heritable: IL-6, IL-12p70, $CD4^+$ cells, IL-13, $CD8^+$, NK cells, IFN-γ and IL-4. However only half of these immunological traits exhibit significant heritabilities: IL-12p70, IL-13, NK cells and IL-4. And of these IL-12p70 and IL-13 have larger site effect contributions to phenotypic variation than a genetic contribution.

Site effects were removed from the analyses by using the HW mice only. With site effects removed six phenotypes were significantly heritable, three the same as the previous analyses, IgG, IgE and IL-12p70 ($P<0.05$). Only IL-13 did not yield a result as the analyses could not converge, likely due to the limited sample size of this study. Three additional phenotypes showed significant heritabilities, IFN-γ, $CD8^+$ and IL-10. While five phenotypes were no longer significantly heritable B cells, NK cells, IL-1b, IL-4 and IL-13.

Phenotype	Heritability		*P*	Site Effect	
B Cells	**0.24**	**(0.16)**	**0.03**	0.08	(0.29)
$CD4^+$	0.34	(0.20)	0.2	~$1x10^{-6}$	(0.41)
$CD8^+$	0.39	(0.23)	0.5	~$2x10^{-6}$	(0.51)
Macrophages	0.12	(0.11)	0.5	~$2x10^{-6}$	(0.98)
NK Cells	**0.41**	**(0.21)**	**0.003**	~$1x10^{-6}$	(0.35)
IFN-γ	0.43	(0.26)	0.5	~$2x10^{-6}$	(0.52)
IL-1b	**0.14**	**(0.11)**	**0.02**	0.67	(0.18)
IL-4	**0.50**	**(0.31)**	**0.001**	~$1x10^{-6}$	(0.50)
IL-6	0.27	(0.22)	0.07	~$1x10^{-6}$	(0.45)
IL-10	0.07	(0.09)	0.1	0.58	(0.24)
IL-12p40	0.13	(0.14)	0.09	0.50	(0.29)
IL-12p70	**0.28**	**(0.17)**	**<0.001**	0.56	(0.25)
IL-13	**0.35**	**(0.20)**	**<0.001**	0.58	(0.23)
IgG	**0.22**	**(0.12)**	**<0.001**	~$1x10^{-6}$	(0.26)
IgE	**0.12**	**(0.08)**	**<0.001**	0.65	(0.18)

Table 3.2. The heritability values of immunological traits, with standard errors (s.e) in wild house mice. Traits in bold have heritabilities which are significantly greater than zero. The site effects are also given, with standard errors (s.e), the value is proportional to the amount of phenotypic variance attributed to differences among sample sites, with a maximum of 1.

Phenotype	Heritability	Standard Error	*P*
B Cells	0.26	0.21	0.096
$CD4^+$	0.19	0.2	0.108
$CD8^+$	**0.36**	**0.2**	**0.012**
Dendritic Cells	0.53	0.21	0.085
Macrophages	<0.001	0.26	0.5
Neutrophils	<0.001	1.26	5
NK Cells	0.22	0.24	0.233
IgG	**0.51**	**0.15**	**0.002**
IgA	**0.23**	**0.14**	**0.009**
IgE	<0.001	0.18	0.5
IFN-γ	**0.65**	**0.22**	**0.018**
IL-1b	0.29	0.33	0.168
IL-4	0.31	0.36	0.233
IL-6	0.13	0.37	0.388
IL-10	**0.88**	**0.14**	**0.002**
IL-12p40	0.22	0.28	0.183
IL-12p70	**1**	**0.09**	**<0.001**
IL-13	-	-	-

Table 3.3. The heritability values of immunological traits, with standard errors (s.e) in wild house mice from the HW sample site only. Traits in bold have heritabilities which are significantly greater than zero.

Discussion

This study is the first to use a marker-based approach to investigate the heritability of measures of immune function in wild house mice. Heritabilities were successfully calculated for 15 immunological phenotypes investigated, with eight of these being significantly different from zero. This study demonstrates the potential use of marker-based approaches to study heritabilities of various quantitative traits in wild populations.

Despite the heritabilities of eight immunological traits being significantly higher than zero the average heritability for these traits was only 0.28, the highest was for concentration of IL4 (0.5) and lowest for concentration of IgE (0.12). These results are considerably lower than those previously reported for similar immunological traits in house mice, pigs and humans (Williams-Blangero *et al.*, 2004; Valdar *et al.*, 2006; Clapperton *et al.*, 2008; Flori *et al.*, 2011; Brodin *et al.*, 2015). In the current study, only cells counts for NK and B cells are significantly heritable, with NK cells having the highest cell count heritability at 0.41 (s.e. 0.21). This heritability is almost double of that previously reported for the percentage of NK cells (0.22), of total lymphocytes, in house mice (Valdar *et al.*, 2006). However, the measure of NK cells in this study and the current study are two different measures of the same cell subset, with scaled cell count used in the current study. When compared to the heritability of total cell counts from humans, where most cell populations have a heritability less than 0.5 (Brodin *et al.*, 2015), the heritability estimates from the current study are substantially greater. Brodin *et al.* (2015) found very low heritability estimates for most counts of various cell populations, with the number of $CD4^+$, NK cells, monocytes, and B cells having heritabilities <0.2 and therefore not detectable under their study design (Brodin *et al.*, 2015). The highest heritability for total cell count measures in this human study was for $CD8^+$ with a value of 0.3, which is similar to the heritability of $CD8^+$ calculated here, 0.39 (s.e. 0.23), however this heritability was not significantly different from zero (Brodin *et al.*, 2015).

Site effects on phenotypic variance were low for cell counts, particularly when compared to those for cytokine and antibody measures, suggesting that cell counts are less affected by environmental and genetic differences that exist among the different sample sub-populations. Cytokine heritabilities also differed considerably from values reported from other studies. In humans *Il12p40* was highly heritable (~1.0) (Brodin *et al.*, 2015) but in the current study the value was only 0.13 (s.e. 0.14) and this was not significantly greater than zero (P= 0.09). IL-4, which had the overall highest heritability of immunological traits in the current study (0.5, s.e. 0.31) and was significantly greater than zero (P= 0.001), has also been found to be significantly heritable in a pedigree-based study of a humans (0.7, s.e. 0.1, P<0.0001) but not heritable in a twin study of humans (0.37, 95% CI -0.15 to 0.89, P>0.05) (Höhler *et al.*, 2005). The differences in heritability values reported between this and previous studies of immunological traits is not surprising as the population, species and sample size in each study are different. Despite the current study reporting significant heritabilities for eight of the studied immunological traits further study is required to determine how applicable to the general population of wild house mice this is.

The heritabilities calculated in the current study are likely to be underestimates of the true heritability values for two reasons. Firstly, some heritability may have been incorporated into the site effect values. The site effect value was estimated in order to remove any confounding effect on phenotypic variation among mice. However, some of this variation may be due to genetic variation as some alleles are more frequent in different sub populations. Secondly for a marker-based approach to accurately estimate the heritability of a trait a high density of markers, or a large amount of LD among markers, is required to capture the effect of causal loci (Visscher *et al.*, 2008). Low amounts of LD were identified for the loci used in the current study, (Chapter 2), and the density of loci was very low compared to GWAS studies in humans or cattle (Thompson-Crispi *et al.*, 2014; Brodin *et al.*, 2015). Therefore, it is highly likely that the effects of multiple causal loci have been missed using the current set of SNP markers, resulting in an underestimation of heritability. Future studies would be advised to use more genetic markers with high LD among them in order to capture all

of the genetic contribution to a traits phenotype and produce a less conservative estimation of heritability.

Upon further consideration of these data it was determined that age likely plays a significant effect on certain measures of immune function. With age there is an increased exposure to pathogens, epigenetic changes and fluctuations in resources, therefore age should be included as a covariate along with sex. However, in the analysis weight was used as a proxy for age in these analyses as previous findings found that age and weight are highly correlated (Abolins *et al.* 2017). Additionally, age was only an estimate, calculated from dried eye lens weight, resulting in some mice having negative values using this method. Therefore, weight was a more complete and reliable measure to include as a covariate. In total six significantly heritable phenotypes were identified, half of which were common to the previous analysis IgG, IgE and IL-12p70. All three of these phenotypes exhibited greater heritabilities in the HW sub-population compared to the analysis of all mice. This suggests that these phenotypes are under strong genetic control which is diluted by site covariance and that age/weight does not significantly affect these traits. Five phenotypes are no longer significantly heritable (B cells, NK cells, IL-1b, IL-4 and IL-13), this may be due to the inclusion of age/weight removing a significant proportion of the variance or the site effects that were present in the initial analysis.

Chapter 4: The Genetic Influence of Immune Function in Wild *Mus musculus domesticus*

Abstract

We know from previous chapters that wild house mice are genetically diverse and that some measures of immune function are significantly heritable, suggesting that genetic variation may play a key role in determining the immune responses of wild house mice. The majority of studies that have examined the effect of specific genetic variants on immune function have looked for associations between the genotype and resistance/susceptibility to infections. Very few studies have sought associations between specific genetic variants and direct immunological measures such as cell counts, antibody titres and serum cytokine concentration. In this study candidate genes were selected based on their function and relation to the immunological measures being recorded. Databases were then used to identify potential polymorphic regions of candidate genes that were then amplified via PCR for a sample of the collected wild house mice and sequenced. From the sequence data SNPs were identified and characterised. A total of 70 SNPs were identified in the exon regions of the candidate genes, with the vast majority of these being novel to this study. The sequencing confirmed the highly polymorphic nature of the H2 locus in house mice and the surprising variation in IL1a. KASP genotyping was then carried out on all of the sampled mice. These data were then used in linear models to seek genetic associations. In total seven significant associations were identified in these house mice between four SNPs and four phenotypes, all of which were novel to this study. These results suggest that wild house mice are excellent models for human populations as they share more environmental and genetic variations that laboratory bred mice but can have laboratory-developed resources applied to them. Wild house mice therefore present a unique research opportunity in the development of the field of immunogenetics.

Introduction

Heritability of immune function reveals to what extent variation in immune function is accounted for by genetic variation, however this does not reveal which genetic variants contribute to this variation. Laboratory studies have revealed the ontology of many genes, including those involved in the immune system and what role they play (UniProt: the universal protein knowledgebase). Additional studies have also investigated the association of specific polymorphisms in these genes to variation in immune function among individuals of the same species. However, the majority of these studies have been carried out with laboratory animals, with genetically homogeneous populations, and focused primarily on the major histocompatibility complex (MHC) (reviewed in Piertney and Oliver, 2006)). Therefore, seeking genetic associations within wild populations between genes and immunological genotypes beyond the MHC may offer novel insights into immune function that are overlooked by "traditional" laboratory-based studies.

There are two predominant methods for conducting a genetic association study; Genome wide association studies (GWAS) and the candidate gene approach. GWAS methods use a large number of genetic markers spread across the genome and is a phenotype-first approach, in which the phenotype of individuals is first characterised, and then potential genetic associations are sought between the phenotype(s) and a range of genetic SNP variants. GWAS is often conducted using a case-control methodology in which a test group is compared to a control group, for example individuals with and without a disease. An odds ratio between the two groups can then be calculated and significance tested using a chi squared test. In order to find significant associations, GWAS requires a well characterised genome with a large number of known SNP loci and a high density of coverage across the genome to account for linkage to causative loci. The application of GWAS to wild populations is therefore limited, particularly if there is limited information about the relatedness of individuals within the population, or the amount of linkage disequilibrium that exists in the population, which may result in synthetic associations being reported.

GWAS approach have been successfully utilised to examine genetic associations pertaining to immune function of mammals (Xavier and Rioux, 2008). However, the limitations of these studies, mentioned above, mean that they are often only performed on livestock, model organisms and humans. Additionally, associations are usually sought to the resistance / susceptibility of hosts to infections rather than specific immunological measures (Xavier and Rioux, 2008). However, there have been relatively few GWAS studies that have examined direct associations to specific immunological measures of immune function such as antibody titres or immune cell counts. A further limitation of conducting GWAS is the inability of the methodology to account for epistasis among genes, therefore there may be false positive reported due to linkage to masking loci (Platt *et al.*, 2010; Ritchie and Van Steen, 2018)

In contrast to a GWAS methodology a candidate gene approach uses *a priori* information about gene ontology to identify genetic associations. Loci are selected based on their predicted effects on a specific phenotype and associations between genotypes and phenotypes are then sought (Amos *et al.*, 2011). Candidate gene approach studies are more applicable to situations where the study species had a limited known genome coverage, where there is a small population size, or where the relatedness of individuals is unknown. As little information about the relatedness or population size of the mouse population sampled was available in this study, a GWAS was not an appropriate method of genetic association to use and therefore a candidate gene approach was used.

Major Histocompatibility Complex (MHC) Variation

The MHC plays a major role in the immune resistance to pathogenic infections in vertebrates and is an interface between the host and pathogens. Immunological genes of the MHC locus are defined as either class I or class II, in mice the MHC is referred to as the H-2 locus, from here on simply H-2. Class I genes are ubiquitously expressed in nucleated cells and present intracellular pathogenic antigens to pathogen recognising cells of the immune system. The expression of class II genes is restricted to that of specific antigen presenting cells (APCs) such as dendritic and B cells. MHC genes are one of the most polymorphic class of genes in vertebrates (Li *et al.*, 2017) and the immunological function of their products is better understood than the majority of immune genes (Radwan *et al.*, 2010).

Conservation programmes have recently become interested in the diversity that organisms possess at MHC loci, which is necessary to provide adequate resistance to infectious disease (Hughes, 1991) in their breeding programmes. Balancing selection by pathogens plays a major role in maintaining genetic/immunological variation (Radwan *et al.*, 2010). This occurs because pathogenic agents are able to rapidly adapt to common host genotypes, to evade the effects of the immune system, but are less likely to adapt rapidly to rare allelic variants resulting in frequency dependent selection of alleles (Borghans *et al.*, 2004). Studies also report that the genes of the MHC experience heterozygote advantage in the resistance to diverse pathogenic infections as heterozygote individuals are able to respond to a larger antigen repertoire (Doherty and Zinkernagel, 1975; Nei and Hughes, 1991). For example, house mice that were heterozygote at the H-2 locus were more resistant to infection with multiple strains of *Salmonella* and one of *Listeria* compared to their homozygote parental lineages (Penn *et al.*, 2002). Further, the excess of reported non-synonymous SNPs compared to synonymous SNPs in extracellular regions of the MHC gene products suggests that there is a positive selection for amino acid substitutions in these functionally important regions of the expressed proteins (Garrigan and Hedrick, 2003; Sommer, 2005; Piertney and Oliver, 2006).

Non-MHC Genetic Variation

Although the effect of variation in genes of the MHC has been well studied in laboratory and wild animals (Sommer, 2005), it would be highly reductionist to only consider the role of the MHC genes in immunological variation among individuals. Up to 5% of genes in the mammalian genome have immune related functions (Trowsdale and Parham, 2004) and the MHC represents only a very small sample of these genes. Genetic associations to pathogen susceptibility have also been shown to occur in many non-MHC genes (Acevedo-Whitehouse and Cunningham, 2006). In fact the majority of risk variants reported from GWAS studies are found in non-MHC genes (Ferreira, 2018).

Among inbred strains of house mice measures of immune function can be highly variable (Sellers *et al.*, 2012). As genetic background is the main variable in laboratory studies it is assumed that genotype significantly affects immune phenotype. Some of the underlying genetic variants among these laboratory mouse strains have been identified. In the innate immune system Toll-like receptors recognise PAMPs on the surface of pathogenic agents. Functional TLR4 is able to detect and respond to challenge with LPS initiating a signalling cascade, however in the strain C3H/HeJ there is a point mutation that renders this strain unresponsive to challenge with LPS (Poltorak *et al.*, 1998). In addition a different point mutation reporting in a nonsense mutation also causes C57BL/10ScCr to not respond to LPS challenge (Poltorak *et al.*, 1998). NK cells are key components of the innate immune system, Killer Cell Lectin-Like Receptors (*Klra*) genes encode for a set of proteins expressed on the surface of NK cells, that recognise and bind to the MHC class 1 receptors which can then initiate the activation or suppression of the NK cell. Among several inbred strains of house mice, the expression of *Klra* genes varies considerably. *Klra15* (*Ly490* and *Ly49L*) are expressed in 129/J mouse strains and *Klra12 (Ly49L)* is expressed in both CBA/J and C3H/He strains, however these alleles are not expressed in the common C57BL/6 strain (Mayer *et al.*, 1980). This difference in expression may be the underlying reason for the increased resistance of tumour induction in 129/J mice compared to C57BL/6 mice (Smith *et al.*, 1973; Mayer *et al.*, 1980)

The adaptive immune system also displays substantial amounts of phenotypic variance among strains of house mice. However, variation in the adaptive immune system is of particular interest to humans for its role in autoimmune diseases such as asthma, arthritis and colitis. House mice can be used to model arthritis by use of K/BxN serum transfer from an arthritic mouse into healthy mice. In response to serum transfer BALB/c mice are high responders in comparison to mice from the SJL strain (Ohmura *et al.*, 2005). This difference in arthritic response is mediated by differential expression of the *Il1b* gene which is substantially higher in the BALB/c mice (Ohmura *et al.*, 2005). Similarly, ulcerative colitis, which is an inflammatory disease of the lower digestive system, shows different levels of susceptibility between inbred house mouse strains. C57BL/6J only develop mild colitis symptoms in response to IL-10 deficiency whereas C3H/HeJBir have a much higher susceptibility of inflammatory bowel disease (Mähler and Leiter, 2002). Together these studies demonstrate that genetic variation plays a significant role in determining the immune responses of house mice

Aims

- Identify genetic variation in candidate genes in wild house mice
- Characterise the polymorphisms reported in the study
- Genotype these polymorphisms on a larger scale
- Seek genetic associations between candidate gene polymorphisms and measures of immune function

Materials and Methods

Locus Specific Genotyping

Candidate Gene Selection

23 genes with protein products that have functional roles in immune state were selected for investigation, as well as one non-immune related gene *MYO1A* (Table 4.1). Gene selection was further guided by the set of immune parameters being measured in this study. The final list of genes cover a range of immunological functions such as cell signalling and pathogen detection, and also represent both the innate and adaptive branches of the mammalian immune system.

Major Histocompatibility Complex

The Major Histocompatibility Complex (H2 in mice): are a class of antigen presenting receptors that are important in the initiation of both innate and adaptive immune responses. Class I molecules present intracellular antigens to cytotoxic ($CD8^+$) T cells triggering an immune response if non-self-antigens are presented. Therefore, polymorphisms in these genes may alter the proliferation and maturity of $CD8^+$ T cells and their downstream effects. Class 2 molecules are expressed by specific antigen presenting cells and present antigens from extracellular pathogens that have been phagocytosed. The antigen-MHC class II complex is recognised by receptors on T-helper cells and used to mediate an appropriate immune response. Therefore, polymorphisms in these genes may result in changes to B and T-cell counts, activation and antibody titres, following stimulation. The MHC genes are highly polymorphic in order to produce variable receptors to bind to more pathogenic antigens and are the most well studied class of immune genes (Sommer, 2005; Acevedo-Whitehouse and Cunningham, 2006). These studies have revealed that the genes comprising the MHC complex are highly variable in their genetic composition. This variation is essential for host organisms to resist pathogenic infections as having a large repertoire of antigen receptors allows for a wider range of antigens to be bound and presented.

Toll like Receptors

Toll like receptors are a class or immunological receptors that are a key component of the innate immune system in found in many animal species. However, unlike the receptors of the MHC toll-like receptors have highly conserved structural regions in order to recognise pathogen-associated molecular patterns (PAMPs). Different TLRs mediate different immune responses to specific antigens. TLR4 recognises liposaccharides, TLR5 recognises flagellin and TLR9 recognise CpG. As these receptors have highly conserved PAMP regions it is expected that any mutation would have significant effects on immune function. However, any non-synonymous polymorphisms are likely to be very rare in wild house mice due to selection against this and can therefore be used as a comparison to MHC receptor genes.

MYO1A

MYO1A is a myosin protein associated with motor function. The gene is not normally expressed in immune cells and therefore was selected to act as a control to compare immune genes against.

Cytokines

Cytokines are immunological messenger molecules, used to coordinate the actions of the immune system. Cytokines have widespread reaching direct and indirect effects and can modify the activation and proliferation of immune cells and affect the expression of immunologically active compound such as antibodies and complement molecules. Therefore, alterations to the expression and function of cytokines, via polymorphisms, can have significant effects on immune function (ref). The following cytokines have been included in order to cover a range of effects on different cell types and antibody effects.

SNP Targeting

Sequence amplification was targeted to predicted SNPs (based on DNA sequence comparisons among strains of inbred house mice) in order to simplify the *de novo* SNP discovery process (Table 4.2). Predicted SNPs were obtained from the MGI database (http://www.informatics.jax.org/). Coding non-synonymous SNPs were prioritised followed by synonymous SNPs and finally SNPs present in non-coding regions or untranslated regions of the candidate gene.

Gene Name	Gene Symbol	Summary of function of gene product
Interleukin 1 A	*Il1a*	Produced by activated macrophages, induces IL-2 release and B-cell maturation and proliferation
Interleukin 1 B	*Il1b*	Pro-inflammatory cytokine, induces neutrophil influx and activation, T-cell activation and cytokine production and B-cell activation and antibody production
Interleukin 2	*Il2*	Released by T-cells, stimulates T-cell proliferation, can stimulate numerous immune cells and is involved in the regulation of immune activities
Interleukin 4	*Il4*	Involved in B-cell activation and induces the expression of class II MHC molecules, enhances secretion of IgG_1 and IgE
Interleukin 6	*Il6*	Pro-inflammatory cytokine and inducer of the acute phase, import for differentiation of B-cells
Interleukin 10	*Il10*	Anti-inflammatory cytokine that inhibits the synthesis of several cytokines, secreted by macrophages and helper T-cells
Interleukin 12B/ p40	*Il12p/p40*	Growth factor for activated T and NK cells, stimulates the production of IFN-γ
Interleukin 12p70	*Il12-p70*	*Il12* subunit that forms a heterodimer with IL12-p40
Interleukin 13	*Il13*	Anti-inflammatory cytokine that inhibits inflammatory cytokine production, important role in regulating inflammatory responses
Interleukin 17 A	*Il17a*	Involved in stimulating the production of pro-inflammatory cytokines
Interleukin 17 F	*Il17b*	Involved in stimulating the production of pro-inflammatory cytokines and involved in the proliferation of peripheral blood mononuclear cells
Interferon gamma	*Ifng*	Activator of macrophages, with immuno-regulatory effects
Tumour necrosis factor	*Tnf*	Produced by macrophages and induces IL-1 secretion
Interleukin 2 receptor gamma	*Il2rg*	Common subunit for a range interleukins
Cluster differentiation 40 ligand	*Cd40lg*	Co-stimulator of T-cell proliferation and cytokine production, particularly IL-4 and IL-10

Toll-like receptor 4	*Tlr4*	Mediates innate immune response to bacterial liposaccharide
Toll-like receptor 5	*Tlr5*	Mediates innate immune responses to bacterial flagellins
Toll-like receptor 9	*Tlr9*	Recognises and mediates immune responses to un-methylated cytidine-phosphate-guanosine (CpG)
Histocompatibility complex Aa	*H2Aa*	Major histocompatibility complex class 2 antigen
Histocompatibility complex Ab	*H2Ab*	Major histocompatibility complex class 2 antigen
Histocompatibility complex Eb	*H2Eb*	Major histocompatibility complex class 2 antigen
Histocompatibility complex K1	*H2K1*	Major histocompatibility complex class 1 antigen
Histocompatibility complex D1	*H2D1*	Major histocompatibility complex class 1 antigen
Unconventional Myosin-Ia	*Myo1a*	Involved in organelle movement within cells, along actin filaments

Table 4.1. A summary of the candidate immunological genes selected for investigation and their immunological functions.

Gene	Chromosome	Exon	No. SNPs	Forward Primer	Reverse Primer	Annealing Temp	PCR Range	Product size
Il1a	2	4	1	CAGATCATGGGTTATGGACTGC	TCCTTCTATGATGCAAGCTATGG	58.10 - 58.86	129306468-129306667	222
Il1b	2	6 - 7	1	TCAAAGCAATGTGCTGGTGC	AGGAGAACCAAGCAACGACA	59.93 - 59.97	129364613-129366065	1453
Il2	3	4	1	GGAGAGCTTTATTTCTTGAAAACAC	TCTGACAACACATTTGAGTGCC	57.14 - 59.38	37120725-37121193	490
Il4	11	1 - 2	0	CGTTGCTGTGAGGACGTTTG	AAACTTAATTGTCTCTCGTCACTG	57.23 - 60.04	53618173-53618647	475
Il6	5	5	1	GTCCTTCCTACCCCAATTTCCA	TCCAAGAAACCATCTGGCTAGG	59.69 - 59.76	30019402-30019833	453
Il10	1	3 - 4	3	ACTTGGGTTGCCAAGCCTTA	TATTAAAATCACTCTTCACCTGCTC	57.15- 59.91	131021333-131022552	1220
Il12a	3	7/8	1	TCTCTGAATCATAATGGCGAGACT	TTTTAAATAAGGGGTGACTGAGTGT	58.39- 59.41	68697895-68698394	500
Il12b	11	6 - 7	1	TGAAGGAGACAGAGGAGGGG	GAACACATGCCCACTTGCTG	59.96 - 60.04	44411101-44412635	1535
Il13	1	1	2	CTGGTCTTGTGTGATGTTGCTC	CAGCCTAGGCCAGCCCAC	59.77 - 62.5	53634494-53634693	200
Il17a	1	3	1	CCTCTGTGATCTGGGAAGCTC	CAGAGTAGGGAGCTAAATTATCCA	57.33- 59.86	20733648-20734449	802
Il17f	1	3	1	TGTGTGCTTCTTCCTTGCCA	CAGACACTCAGGCTGCATCA	60.04 - 60.11	20777324-2.777953	630
Ifng	10	2 - 3	0	TCAAGTGGCATAGATGTGGAAGA	CAAACTTGGCAATACTCATGAATGC	59.48 - 59.71	118442463-118442793	527
Tnf	17	1	3	TCATCCCTTTGGGGACCGAT	CTCAGCGAGGACAGCAAGG	60.45 - 60.62	35201656-35201995	340
Il2rg	X	8	0	GGTAGAAAAAGGGAGGGAGAATCC	CTGCAGCCAGACTACAGTGA	59.39 - 60.39	101264439-101265001	563
Cd40lg	X	5	1	ACAGTGGGCCAAGAAAGGAT	GGAGCCCAGGTCAACCATAA	59.22 - 59.38	57223215-57224026	812
Tlr4	4	3	2	AGTGGCCCTACCAAGTCTCA	GCGGGGCACTCCTTCTTCTA	60.18 - 61.61	66840072-66841071	1000
Tlr5	1	4	1	CAGACGGCAGGATAGCCTTT	TGCAGAGGCTCGAGTTCATC	59.82 - 59.83	182973251-182973739	508
Tlr9	9	2	4	ATGTGGCCAAAAGTCCCTCC	ATACCGTTGCCGCTGAAGTC	60.25 - 60.74	106224312-106225324	1032
H2Aa	17	2 - 3	14	TGTTTCAGAACCGGCTCCTC	ACCACGTAGGCACCTATGGTA	59.97 - 60.34	34283580-34284533	954
H2Ab	17	3 - 4	5	AGCCCAATGTCGTCATCTCC	TGTGACGGATGAAAAGGCCA	59.82 - 59.89	34267339 - 34268002	683
H2Eb	17	3	5	GCCTACGGTGACTGTGTACC	GCTGGGATGCTCCACCTG	60.11 - 60.12	34314174-34314406	250
H2K1	17	2 - 3	10	GCGTTCCCGTTCTTCAGGTA	TGAGGTATTTCGTCACCGCC	60.04 - 60.11	33999288-33999972	704
H2D1	17	2 - 3	14	CCCACACTCGATGCGGTATT	GCGTTCCCGTTCTTCAGGTA	60.04 - 60.18	35263380-35264078	718
Myo1a	10	8 - 10	11	AGGAGTATGTTATCCCCAACCA	TGTGACATTCAGCACCGTGA	58.26 - 59.89	127710010-127710993	984

Table 4.2 A summary of the PCR primers designed to amplify target regions of the candidate genes. The chromosome of each candidate gene is listed as well as the targeted exon of the PCR primer and the number of SNPs previously reported among laboratory strains of house mice. For IL-12p70 the targeted exon is reported as 7 in some genome assemblies and 8 in others. The sequence of these primers and their annealing conditions, the genomic range of these PCR products and the expected PCR product length are all shown. *H2*K1 and *H2*D1 failed to amplify using the above PCR primers so were not included in further genotyping assays.

PCR Primer Design

PCR primers were designed to amplify targeted gene fragments of up to 1.5kb in length from the extracted wild mouse genomic DNA (Table 4.2), see chapter 2 for DNA extraction protocol. Primers were designed based on the genomic sequence of C57BL/6 laboratory mice. The National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/) was used to obtain the targeted genomic DNA sequences for each gene. Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was then used to design the primers using specified parameters. Primer specifications were set to 60^{o}C for T_m (melting temperature) and the target nucleotide length for primers was 20. The product length for each primer set was varied depending on the respective gene and care was taken to ensure that both primers were situated in the exons of their respective genes to ensure amplification due to a higher probability of sequence variation in intronic or non-coding regions inhibiting the binding of complementary primers designed from C57BL/6 sequences. The final primer sets had a T_m range of 57.14 to 62.5^{o}C and nucleotide length of 200 to 1535. The primers were then ordered from ThermoFisher Scientific [Thermo Fisher Scientific UK] and re-suspended with pure water to give a working concentration of 10μM.

PCR - sample selection

De novo SNP discovery was carried out on a subset of mice selected to proportionally represent each of the sample sites in the study (Table 4.3). The DNA from these sample mice was used in PCR reactions with each of the candidate gene primer sets.

Mouse I.D	Site
17	Barrow Mill Farm
20	Barrow Mill Farm
37	Green Lanes Farm
53	Hyatt's Wood Farm
64	Parsonage Farm
107	J.B. Equestrian
112	Hyatt's Wood Farm
123	Westrip Farm
137	Pixie Hall
157	Sperrings Farm
244	Hyatt's Wood Farm
251	Barrow Mill Farm
265	London Underground
267	London Underground
282	Hyatt's Wood Farm
290	Woefuldane Farm
309	Hyatt's Wood Farm
317	Pixie Hall
327	Pixie Hall
351	Skokholm Island
354	Skokholm Island
372	Hyatt's Wood Farm
379	Hyatt's Wood Farm
382	Hyatt's Wood Farm
403	J.B. Equestrian

Table 4.3. The DNA from the mice was used in single read Sanger DNA sequencing. The mice were selected to proportionally represent each of the sample populations of the study in order to capture variation in each.

PCR - Conditions

Targeted gene fragments were amplified through polymerase chain reaction (PCR). The PCR annealing temperature (T_a) was initially optimised by carrying out a gradient PCR experiment for each primer set. A 25 μL reaction was run for each primer set across a temperature gradient ranging from 55°C to 65°C, using 1μL of genomic C57BL/6 DNA [Promega], 2μL of forward primer (10μM), 2μL of reverse primer (10μM), 2.5μL (10 units) of Platinum Taq High fidelity Buffer [Life Technologies], 0.5μL of DNTPs (10mM each) [ThermoFisher Scientific UK], 1μL MgSO4 (mM) [Life Technologies] and 16μL of pure H_2O [Sigma Aldrich]. The following PCR program was used in these reactions: denaturation at 94°C for 30s, annealing (gradient temperatures) 15s, elongation 68°C time depending on product length (30s per kb) and a final elongation at 68°C for 10 minutes. PCR products were then run out on a 1% w/v agarose gel stained with ethidium bromide and imaged under U.V. light. Once the optimum annealing temperature was identified for each primer set PCR with these primers was carried out on the selected sample of wild mice in 50μl reactions, doubling all reagents stated above. The PCR products were again run on a 1% w/v agarose gel with a 1kb DNA ladder [Thermo Fisher Scientific UK] to confirm correct amplification based on the product lengths. Successful PCR products were purified using then GeneJet PCR clean-up kit [Thermo Fisher Scientific UK]. The PCR clean-up was carried out according to the manufacturer's instructions using 40μl of PCR product for each sample.

Sanger Sequencing

Single-read sequencing of the purified PCR products was carried out by Eurofins Genomics [Eurofins Scientific]. For each gene of interest 15μl of purified PCR product was premixed with 2μl (10μM) forward primer. Samples that failed to sequence satisfactorily for a particular candidate gene were re-sequenced with 2μl of the reverse PCR primer (10μM) (Table 4.4).

Gene	No. Individuals Sequenced	Exon Sequence Length	No. SNPs in Exons	Intron Sequence Length	No. SNPs in Introns	Bp per SNP (Exons)	Bp per SNP (Introns)
Il1a	23	157	1	-	-	157	-
Il1b	20	572	10	201	10	57.2	20.1
Il2	22	371	-	-	-	-	-
Il4	22	162	1	191	-	162	-
Il6	22	388	1	-	-	338	-
Il10	20	99	-	667	7	-	95.3
Il12a	24	435	-	-	-	-	-
Il12b	17	-	-	783	1	-	783
Il13	22	138	2	-	-	69	-
Il17a	15	732	5	-	-	146.4	-
Il17f	21	564	3	-	-	188	-
Ifng	22	172	0	45	-	-	-
Tnf	19	272	3	-	-	90.7	-
Il2rg	17	507	0	-	-	-	-
Cd40lg	17	636	0	-	-	-	-
Tlr4	21	744	5	-	-	148.8	-
Tlr5	21	443	3	-	-	147.7	-
Tlr9	21	828	1	-	-	828	-
H2Aa	13	209	11	290	26	19	11.2
H2Ab	22	276	14	305	20	19.7	15.3
H2Eb	23	187	7	-	-	26.7	-
Myo1a	21	252	2	539	10	126	53.9

Table 4.4. A summary of the single read Sanger DNA sequencing of PCR products from the pre-selected mouse samples. Genes that are highlighted failed to identify any SNP loci in the regions that were sequenced among the sampled mice, for these genes the previously predicted SNP loci (based on the MGI database) were used for KASP genotyping. Dashes show where no data was obtained. For comparisons of SNP diversity among the sequences the number of base pairs (Bp) per SNP was calculated for both exon and intron sequences.

SNP Discovery

Individual FASTA files from successful sequencing were compiled into a single FASTA file for each candidate gene. The sequences within each of the compiled FASTA files were then aligned using ClustralW (Thompson, Higgins and Gibson 1994) within MEGA6 [Molecular Evolutionary Genetics Analysis] (Tamaura *et al.* 2013) using the default parameters. SNPs in each candidate gene were then manually identified from the aligned sequences (Tables 4.5 and 4.6). The identified SNPs were further confirmed using NCBI BLAST [National Centre for Biotechnology Information] and the MGI [Mouse Genome Informatics (http://www.informatics.jax.org/)] database and the possible changes to the amino acid sequences were identified using the IGV genome viewer with the mouse construct 10 [Integrative Genomics Viewer].

Gene	SNP	Seq Site	Genomic Site	No.Mice	Codon(AA) change	Syn/ Non-syn
Il1a	**C/T**	**93**	**129306612**	**12/23**	**-**	**Syn**
Il1b	**A/C**	**21**	**129364687**	**12/20**	**UTR**	**-**
	G/A	39	129364705	1/20	UTR	-
	C/A	177	129364843	1/20	UTR	-
	C/T	252	129364918	1/20	UTR	-
	A/C	306	129364972	1/20	UTR	-
	A/C	320	129364986	1/20	UTR	-
	C/A	**392**	**129365058**	**1/20**	**Asp>Tyr**	**Non-syn**
	C/T	432	129365098	1/20	Phe>Phe	Syn
	G/A	435	129365101	1/20	Val>Val	Syn
	T/C	531	129365197	1/20	Phe>Phe	Syn
Il4	**C/G**	**220**	**53618503**	**1/22**	**Gly>Ala**	**Non-syn**
Il6	**C/T**	**92**	**30019545**	**4/22**	**Thr>Thr**	**Syn**
Il13	T/C	21	53634563	2/22	-	Syn
	C/T	**27**	**53634569**	**2/23**	**-**	**Syn**
Il17a	**G/A**	**98**	**20733797**	**2/15**	**Glu>Lys**	**Non-syn**
	A/G	165	20733864	2/15	UTR	-
	C/T	218	20733917	6/9	UTR	-
	G/A	**392**	**20734091**	**7/15**	**UTR**	**-**
	A/G	498	20734197	3/10	UTR	-
Il17f	A/T	68	20777440	1/21	UTR	-
	T/C	**174**	**20777546**	**11/19**	**UTR**	**-**
	T/C	419	20777791	1/19	Lys>Lys	-
Tnf	**G/T**	**114**	**35201817**	**4/19**	**Ala>Ser**	**Non-syn**
	C/A	227	35201930	5/19	UTR	-
	G/A	236	35201939	2/19	UTR	-
Tlr4	C/T	555	66840696	20/21	Asn>Asn	Syn
	C/A	580	68840711	1/21	His>Asnμ	Non-syn
	A/G	581	68840712	1/21	His>Argμ	Non-syn
	A/C	**609**	**68840750**	**20/21**	**Glu>Asp**	**Non-syn**
	T/C	682	68840823	1/21	Leu>Leu	Syn
Tlr5	T/C	15	182973315	1/21	Thr>Thr	Syn
	G/A	**255**	**182973555**	**1/22**	**Leu>Leu**	**Syn**
	A/G	267	182973567	1/23	Gly>Gly	Syn

Tlr9	**C/T**	**175**	**106224547**	**1/22**	**Arg>Cys**	**Non-syn**
H2Aa	G/A	36	34283683	2/11	Tyr>His	Non-syn
	A/C	40	34283687	1/11	Arg>Arg	Syn
	G/A	49	34283696	1/11	Phe>Phe	Syn
	G/A	**51**	**34283698**	**2/11**	**Phe>Leu**	**Non-syn**
	C/T	61	34283708	2/11	Glu>Glu	Syn
	G/A	70	34283717	1/11	Gly>Gly	Syn
	T/C	78	34283725	1/11	Thr>Ala	Non-syn
	G/A	88	34283735	3/11	Ser>Ser	Syn
	T/A	121	34283768	11/12	Pro>Pro	Syn
	G/T	145	34283792	4/12	Ile>Ile	Syn
	A/G	148	34283795	5/12	Leu>Leu	Syn
	G/A	178	34283825	2/13	Ser>Ser	Syn
H2Ab	C/T	13	34267403	5/22	Cys>Cys	Syn
	A/G	16	34267406	10/22	Ser>Ser	Syn
	A/G	34	34267424	5/22	Pro>Pro	Syn
	C/A	59	34267449	10/22	Arg>Arg	Syn
	G/A	**79**	**34267459**	**10/23**	**Gln>Arg**	**Non-syn**
	C/T	156	34267536	3/23	Glu>Lys	Non-syn
	A/G	159	34267539	7/23	Met>Val	Non-syn
	T/G	160	34267540	10/23	Met>Arg	Non-syn
	A/G	162	34267542	10/23	Thr>Ala	Non-syn
	C/T	181	34267561	1/23	Val>Ala	Non-syn
	C/T	184	34267564	10/23	Tyr>Ser	-
	C/G	196	34267576	2/23	-	-
	C/T	543	34267923	3/22	-	-
	C/T	553	34267933	1/22	Ser>Ser	Syn
H2Eb	**G/A**	**21**	**34314245**	**3/23**	**Ser>Asn**	**Non-syn**
	A/G	45	34314269	1/23	Glu>Gly	Non-syn
	T/C	64	34314288	1/23	Asn>Asn	Syn
	A/G	77	34314301	2/23	Lys>Glu	Non-syn
	A/C	133	34314357	11/23	Thr>Thr	Syn
	A/G	149	34314373	2/23	Thr>Ala	Non-syn
	A/G	161	34314385	2/24	Ser>Gly	Non-syn
MYO1A	G/A	308	127710379	1/21	Ala>Ser	Syn
	C/T	**366**	**127710437**	**2/21**	**Leu>Leu**	**Syn**

Table 4.5. Summary of SNP discovery within exon regions of candidate genes. SNPs in bold were selected for KASP genotyping based on the type of SNP and the frequency of these SNPs among the sampled mice. Exon regions *for Il2, Il10, Il12a, Il12b, I*□□□, and *Cd40lg* contained no SNP loci among the sampled wild mice. SNPs found within describes coding regions were denoted as synonymous or non-synonymous, with non-synonymous amino acid changes resolved. Where possible SNP loci in exon regions were selected for KASP sequencing, with priority placed on bases coding for non-synonymous SNPs, followed by synonymous SNPs and SNPs in the untranslated regions (UTR). For *Il17a* two SNPs were selected for KASP genotyping the first, from here on referred to as *Il17a_N*, was found to be a SNP coding for a non-synonymous substitution but only present in two sampled wild mice and the second, referred to from here on as *Il17a_U*, located in the UTR of the gene was found in 11 sampled mice.

Gene	SNP	Seq Site	Genomic Site	No. Mice
Il1b	G/A	606	129365272	1/20
	T/G	630	129365296	1/20
	G/C	644	129365310	1/20
	C/T	651	129365317	8/19
	A/G	667	129365333	8/19
	G/A	674	129365340	10/19
	A/T	695	129365361	8/19
	A/G	697	129365363	1/19
	C/T	739	129365405	8/19
	C/A	767	129365433	1/19
Il10	T/C	138	131021519	4/20
	T/C	**139**	**131021520**	**4/20**
	C/T	323	131021704	3/20
	C/T	383	131021764	1/20
	G/A	500	131021881	3/20
	G/A	540	131021921	1/20
	T/G	628	131022009	1/19
Il12p70	**G/A**	**749**	**44411954**	**4/15**
H2Aa	A/T	214	34283861	1/13
	G/C	236	34283883	1/14
	C/G	240	34283887	2/13
	C/T	246	34283893	3/13
	T/C	264	34283905	3/11
	T/C	269	34283910	3/13
	T/A	279	34283920	4/13
	T/C	305	34283946	3/13
	T/C	307	34283948	3/14
	C/T	317	34283958	4/13
	T/C	327	34283968	3/13
	G/A	328	34283969	4/13
	A/T	329	34283970	1/13
	T/C	330	34283971	2/13
	C/T	337	34283978	3/13
	C/T	342	34283983	1/13
	C/T	349	34283990	2/13
	T/C	360	34284001	4/13

	G/A	361	34284002	4/14
	C/T	379	34284020	7/13
	G/A	380	34284021	2/13
	C/T	438	34284079	1/13
	C/T	442	34284083	2/13
	C/T	460	34284101	1/13
	C/T	463	34284104	3/13
	G/A	480	34284121	5/13
H2Ab	A/C	246	34267626	10/22
	G/A	274	34267654	1/22
	G/C	282	34267662	5/22
	T/A	289	34267669	5/22
	T/C	291	34267671	5/22
	T/C	327	34267707	21/22
	C/G	340	34267720	4/22
	A/G	367	34267747	18/22
	C/T	381	34267761	2/22
	T/C	388	34267768	10/22
	G/A	405	34267785	10/22
	A/G	438	34267818	17/22
	T/C	449	34267829	7/22
	G/A	466	34267846	7/22
	C/T	473	34267853	18/22
	A/T	478	34267858	1/22
	G/A	489	34267869	5/22
	C/T	517	34267897	8/22
	C/T	527	34267907	2/22
Myo1a	C/T	72	127710143	5/21
	C/T	254	127710325	16/20
	C/T	467	127710538	2/19
	T/A	504	127710575	2/19
	T/G	**546**	**127710617**	**12/19**
	C/G	551	127710622	1/19
	C/T	614	127710663	2/19

C/T	636	127710685	2/19
C/T	726	127710775	13/19
T/C	775	127711824	14/16

Table 4.6 Summary of the SNPs identified in the intron regions of candidate genes sequenced from the pre-selected sampled wild house mice. The SNPs shown in bold were selected for KASP genotyping. SNPs were selected from intron regions for KASP genotyping only if no SNPs were identified in the exon region of the same gene. For MYO1A a SNP from both the intron and exon region of the gene were selected for KASP genotyping with these SNPs acting as controls in later analyses.

KASP Genotyping

Following SNP discovery, DNA from mouse samples were sent to LGC Genomics. The 30nt flanking sequences from either side of the SNP were also sent with the samples to allow for primers to be designed for the genotyping assays. 50μL of DNA (between 5ng/μL and 100ng/μL), suspended in TE buffer, was sent for each mouse sample to LGC genomics [LGC Genomics] in 96 well plates.

Following genotyping data was received in an Excel format. Anomalous data were removed from future analyses, this included mice with a high proportion of errors, loci with high a proportion of errors and loci that were highly monomorphic (Table 4.7).

Immunological Data

Immune cells were extracted from the spleens of culled mice and counted using flow cytometry (Abolins *et al.*, 2018). Antibody titres were taken from the blood serum of the house mice following cardiac puncture (method in Abolins *et al.*, 2018).

Analysis

H2-Locus Heterozygosity

The number of viral / bacterial infections the sampled wild house mice possessed was compared to their level of heterozygosity across the three *H2*-locus gene; 0,1,2 or 3. An ANOVA test was carried out in R via RStudio (R Core Team, 2013; RSTudio Team 2015) to determine whether heterozygosity significantly affected the number of infections mice possessed. This first included mice that were heterozygous at all three loci, however the number of individuals was only three, so this analysis was re-run after removing this category. Finally, a two-tailed, uneven-sample t-test was carried out in Excel (Microsoft Excel 2016) comparing the number of infections possessed by homozygous individuals to heterozygous individuals.

Genetic Associations

R (v. 2.15.3) was used was used to construct linear models (lm function) in Rstudio (RStudio 2015) for each of the immunological phenotypes; scaled cell counts and antibody titres. Linear models needed to be constructed for each of the different data sets (gene-of-interest) for each immune phenotype as the number of individuals varied. This resulted in occasions where the minimal model generated varied within a specific immune phenotype.

The Akaike Information Criterion (AIC) index (Akaike 1973) value of each model was compared to the previous model to select the model with the best fit while penalising for overfitting. Only factors that reduced the AIC value compared to the previous model were retained in the sequential models.

For a given gene-of-interest, linear models were built by sequentially adding factors that may influence the immune phenotype. Factors were added in the following order: site, sex, season, age, gene-of-interest. Additionally, the interaction of gene-of-interest by each fixed factor was added. The gene-of-interest factor was added in sequence regardless of a decrease in AIC value as this was the independent variable of this study. Once the model with the lowest AIC, and therefore the best fit was identified, any more complex models were discarded.

In order to produce p-values and F-values to test for significance an ANOVA was used. Models were compared to their previous, less complex, model using an ANOVA test to determine whether the newly added factor or interaction added to the model significantly affected the phenotype. Finally, when an ANOVA test identified a gene-of-interest having a significant effect on an immune phenotype a Tukey's post-hoc test was carried out to determine the specific effect of each genotype on the immune phenotype.

Results

Genetic Variation in Loci of Immune function

Considerable genetic variation was recorded in the immune gene loci, with a total of 144 SNPs identified across all of the loci and an average of 40.8 bp per SNP in introns and 116.3 bp per SNP in exons. The average minor allele frequency among the immune loci was 0.12 compared to 0.23 for the non-immune loci genotyped in chapter 2. However, the non-immune loci that were selected for the Illumina Golden-Gate Genotyping chip were selected for their known variability among inbred strains of house mice and therefore direct comparisons cannot be drawn.

H-2 Polymorphisms

Sections of five genes in the H-2 locus of mice were sequenced from between 11 and 23 wild house mice, class I genes; *H2-K1* and *H2-D1,* and class II genes; *H2-Aa1, H2-Ab1* and *H2-Eb1.* Based on the MGI database [Mouse Genome Informatics] there are around 25 SNP loci predicted in the *H2-K1* and *H2-D1* genes. However, the both sets of PCR primers designed to for the class 1 genes failed to accurately amplify their targeted regions, therefore no further sequencing or genotyping occurred in the wild house mice for these two genes. In comparison there were 11 predicted SNPs in the *H2-Aa1, H2-Ab1* and *H2-Eb1* genes (1, 5 and 5 respectively) based on the MGI database [Mouse Genome Informatics]. The PCR primers for these three genes did successfully amplify the targeted regions, which were also successfully sequenced, and SNPs identified.

From the sequencing of the PCR products of *H2-Aa1, H2-Ab1* and *H2-Eb1* the average exon sequence length was 224 bases the sequences for *H2-Aa* and *H2-Ab* also contained sequences that were within the introns of these genes, an average 298 bases. It should be noted that that while over 20 mice were successfully sequenced for *H2-Ab* and *H2-Eb* only 13 mice were successfully sequenced for *H2-Aa*, (Table 4.5). Following the alignment of the sequences from the sample of wild mice *H2-Aa* and *H2-Ab* were found to contain the highest number of SNPs in exon regions, 12 and 14 respectively, among the sequenced genes, additionally these genes possessed the

most SNPs in intron regions (26 and 20 respectively). For *H2-Aa* the sequenced region of the gene contained one SNP that had previously been reported among laboratory house mice on the MGI database, this SNP was confirmed to be present among 3/11 wild house mice. However, this locus was not selected for further genotyping assays as it was not known to be non-synonymous.

Among the three *H2* loci examined in this study 77 novel SNP loci were discovered; 38 for *H2-Aa*, 34 for *H2-Ab* and 7 for *H2-Eb*. The length of the PCR product sequenced for these genes was relatively small with a total of 499 bases for *H2-Aa*, 581 bases for *H2-Ab* and 187 bases in *H2*-Eb. Despite these short sequence lengths *H2-Aa* and *H2-Ab* contained the highest concentrations of SNP loci as exhibited by the average number of bases per SNP for each gene; 13.1 for *H2-Aa* and 17.1 bases for *H2-Ab*. The closest non-*H2* gene in terms of number of average number of bases per SNP was found in *Il1b* at 38.7 and although only 187 bases of the *H2-Eb* gene were successfully sequenced there were 7 SNPs identified in this sequence, giving an average number of bases per SNP figure of 26.7. In total, 33 SNPs were identified in the exons of the three *H2*-locus genes, in comparison the previously reported SNPs (from the MGI database) were located in the intron or non-coding regions of these genes. Of the 33 SNPs present in exons, 16 were found to code for non-synonymous changes, resulting in the substitution of amino acids. No non-synonymous SNPs have been previously recorded on the MGI database for these genes compared to the 16 that were identified in this study. In general, the frequency of SNPs in introns was much higher than for SNPs in exons, average number of bases per SNPs 57.1 and 154.7 respectively.

Non- H2 Polymorphisms

Following sequencing and alignment of sampled DNA sequences five non-*H2* genes were monomorphic in both intron and exon regions: *Il2*, *Il12p40*, *Ifng*, *Il2rg* and *CD40lg*. Therefore, genotyping of these genes utilised predicted SNPs from the MGI database. A total of 65 SNPs were identified in the sequenced non-*H2* genes regions, 37 SNPs in exons and 28 SNPs in introns. Among the non-*H2* genes studied *Il1b* was the most polymorphic with 20 SNP loci identified, 10 in the intron regions and 10 and an average number of bases per SNP at 57.2 and 20.1 in exons and introns

respectively. In comparison the remaining polymorphic non-*H2* genes had an average number of bases per SNP of 218.3 for exons and 310.7 for introns. There were far fewer non-synonymous SNPs among the non-*H2* loci, 6 compared to the 16 among the *H2*-locus genes.

Genotyping Summary

Allele Frequencies

In total 25 loci were genotyped in 22 candidate genes (*Il1b*, *Il17a* and *Myo1a* were each genotyped at two loci) based on the SNP discovery carried out previously. Of the genotyped loci 9 were found to be monomorphic (*Tlr4*, *Il1b_N*, *Il2*, *Il4*, *Il12p40 Il12p70*, *Ifng*, *Il2rg* and *Cd40lg*) among the genotyped wild house mice. In the remaining 16 loci only half had a minor allele frequency that exceeded 0.1 (Table 4.7) and could therefore be used for genetic association analyses (*Il1a, Il1b_U, Il17a_U, Il17f, H2Aa, H2Ab, H2Eb* and *Myo1a_2*).

H2 Heterozygosity

The ANOVA test compared the number of infections wild mice possessed (Abolins *et al.*, 2018) in three different categories of heterozygosity (0,1,2) combining the three H-2 loci genotyped. A fourth category, those that were heterozygote at each locus, was not included due to the limited number of samples. The ANOVA test showed that the number of infections did not significantly change with increase in heterozygosity at these loci ($F=2.70$, $df=2$, $p=0.068$). However, a one tailed t-test did show that possessing heterozygosity in any of the *H2*-loci does confer a significant decrease in the number of infections a mouse possesses ($p=0.01$). Further investigation found that being heterozygous at a single *H2*-locus does not confer a significant resistance to infections ($p=0.14$) but heterozygosity at two *H2*-loci was significant in reducing the number of infections ($p=0.009$).

	Il1a	*Il1b*	*Il6*	*Il10*	*Il13*	*Il17a_N*	*Il17a_U*	*Il17f*	*Tnf*	*Tlr5*	*Tlr9*	*H2Aa*	*H2Ab*	*H2Eb*	*Myo1a_2*	*Myo1a_1*
BM	0.23	0.06	0.26	0.03	0.24	0.00	0.01	0.27	0.03	0.29	0.00	0.03	0.30	0.00	0.44	0.00
GL	0.08	0.00	0.00	0.20	0.00	0.00	0.40	0.50	0.00	0.00	0.00	0.13	0.02	0.03	0.17	0.00
HW	0.47	0.47	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.4	0.40	0.13	0.00
JB	0.41	0.01	0.00	0.00	0.00	0.01	0.03	0.39	0.00	0.00	0.23	0.00	0.04	0.04	0.04	0.00
LU	0.03	0.03	0.00	0.00	0.07	0.00	0.14	0.17	0.46	0.00	0.00	0.00	0.07	0.00	0.03	0.00
PF	0.00	0.00	0.19	0.13	0.00	0.00	0.00	0.00	0.44	0.13	0.00	0.00	0.44	0.00	0.32	0.00
PH	0.00	0.13	0.00	0.00	0.00	0.00	0.41	0.16	0.00	0.00	0.00	0.28	0.23	0.04	0.44	0.00
SK	0.02	0.02	0.00	0.04	0.00	0.13	0.04	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48
SP	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ST	0.50	0.25	0.00	0.38	0.17	0.00	0.00	0.17	0.00	0.17	0.00	0.17	0.33	0.27	0.42	0.00
WF	0.15	0.00	0.19	0.00	0.00	0.00	0.00	0.42	0.00	0.00	0.00	0.27	0.33	0.00	0.17	0.00
WT	0.25	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.06	0.00	0.00	0.00
TOTAL	0.46	0.39	0.1	0.09	0.06	0.09	0.45	0.35	0.07	0.06	0.05	0.14	0.43	0.22	0.34	0.06

Table 4.7 This table shows the minor allele frequencies for each of the candidate genes that were genotyped by LGC genomics. *Il2*, *Il12p40, Il12p70, Ifng, Il2rg, Cd40lg* and *Tlr4* were found to be completely monomorphic and were removed from further analyses. SNP loci that possessed a minor allele frequency of less than 0.1 were also discounted from further statistical analyses as they would provide a very limited pool of data. SNP loci with a MAf >0.1 are highlighted in grey and were used in subsequent statistical analyses.

Genetic Associations

The focus of this study was to identify significant genetic associations of SNPs in immune related genes to direct immunological measures of immune function. Therefore, significant non-genetic associations will only be briefly discussed, but were important in building minimal linear-models to explain phenotypic variance. For each immune phenotype a minimal model was built to account for possible confounding factors that contribute to phenotypic diversity prior to the addition of genetic factor, termed gene-of-interest, being added. Minimal models occasionally varied among loci for a specific phenotype due to the differences in the number of mice used in the analysis. The linear models that were constructed for each immune phenotype were analysed using an ANOVA test and factors that significantly influenced the immune phenotype were used to form a minimal model for that phenotype (Tables 4.8 and 4.9). Sex was not a significant factor in any of the models constructed and was not included in any of the minimal models. Apart from the number $CD8^+$ T cells all phenotypes had at least 2 factors that significantly influenced them. Site was a significant factor for all phenotypes except the number of $CD8^+$ T and Dendritic Cells (DCs). Season was a significant factor for all phenotypes but number pf $CD8^+$ T and Neutrophil cells and age for all except the concentration of IgA and the number of B cells.

Significant Associations: Cells

B Cells

Site and Season were the only significant non-genetic factors associated with B cell counts (see table 4.8). The minimal non-genetic factor model was B Cells ~ Site + Season. *H2Eb* and *Il1b* genotypes were both significantly associated with variation in B cell counts (p= 0.045 and p= 0.035 respectively). Additionally, site by gene-of-interest interactions were also significant for *H2Ab* and *Myo1a*; B Cells ~ Site + Season + Site: *H2Ab* and B Cells ~ Site + Season + Site: *Myo1a*.

$CD4^+$ T Cells

No significant genetic associations were found for $CD4^+$ T cell counts. Similarly, to B cells, Site and Season were both significant non-genetic factors, however in addition Age was a significant factor contributing to phenotypic variance (see Table 4.8). The minimum model constructed was $CD4^+$ T Cells ~ Site + Season + Age.

$CD8^+$ T Cells

Age was the only factor to be significantly associated with variation in $CD8^+$ T cells (see Table 4.8). Therefore, the minimum model constructed for this phenotype was $CD8^+$ T Cells ~ Age.

Macrophages

Site, Season and Age were all significantly associated with variation in macrophage cell counts (see Table 4.8), giving an overall minimum model of Macrophages ~ Site + Season + Age. In addition, the genetic factor of *H2*Aa was significantly associated (p= 0.048).

Neutrophils

Site and Age were significantly associated with neutrophil cell count variation giving the minimum model of Neutrophils ~ Site + Age (see table 4.8). No significant associations were identified for any of the genetic factors.

Natural Killer Cells

Site, Season and Age were all significantly associated with natural killer cell counts with the minimum model produced being Natural Killer Cells ~ Site + Season + Age (see table 4.8). No genetic factors were significantly associated.

Dendritic Cells

Season and Age were significant factors in all data sets for dendritic cells and, in addition, for the IL-1b genetic data set site was also a significant factor (see table 4.8). Therefore, the minimum model for all but the Il1b data set was Dendritic Cells ~ Season + Age and for the Il1b data set Dendritic Cells ~ Site + Season + Age. So significant associations were found for any of the genetic factors added.

Significant Associations: Antibodies

IgA

Site was the only non-genetic factor to significantly contribute to the concentration of faecal IgA for all but one data set. The minimal model was therefore IgA ~ Site. However, for the IL-17f genetic data set the linear model IgA~ Site + Season + Age did significantly contribute to phenotypic variance (see table 4.8).

IgE

Site, Season and Age all significantly associated with variation in IgE antibody titres. Sex was not significantly associated but was still included in the model as it lowered the AIC value of model. Two immune loci showed significant associations, IL-1a and IL-1b (p<0.001 for both). When the genotypes of these two loci were combined a significant association was also found (p=0.001). A site by gene-of-interest association was also found (p<0.001).

IgG

Site, season and age were significantly associated with IgG antibody titres; therefore, the minimum model was site + season + age. For the genetic terms both *H2*Ab and *H2*Eb genotypes were significantly associated (p=0.03 and 0.015 respectively).

Immune Phenotype	Terms	d.f.	*F-Value*	*p-value*
B Cells				
	Site + Season + Site x *H2Ab*	21	2.23	0.002
	Site + Season + *H2Eb*	2	3.13	0.045
	Site + Season + *Il1b*	2	3.4	0.035
	Site + Season + Site x *Myo1a*	20	1.97	0.009
Macrophages				
	Site + Season + Age + *H2Aa*	2	3.07	0.048
IgE				
	Site + Sex + Season + Age + Site x *H2Aa*	5	3.08	<0.001
	Site + Sex + Season + Age + *Il1a*	2	8.5	<0.001
	Site + Sex + Season + Age + *Il1b*	2	6.8	0.0013
	Site + Sex + Season + Age + *Il1*	5	4.08	0.001
IgG				
	Site + Season + Age + *H2Ab*	2	3.41	0.03
	Site + Season + Age + *H2Eb*	2	4.28	0.015
IgA				
	Site + Site x *Myo1a*	13	2.35	0.0056

Table 4.8. Linear models of all significant genetic associations

Genotype Effects

A Tukey's post-hoc test was carried out for each linear models where the gene-of-interest was a significant factor (Table 4.10). The tests revealed the significant variations between genotypes for each locus at the 0.05 level of significance. From this the mode of inheritance of each allele can be inferred.

Phenotype	Gene-of-Interest	Genotype	Log_{10} Mean	Scaled Cell Counts / Antibody Titres	Significance
B Cells	*H2Eb*	AA	7.02	10471285.5	a
		AG	7.01	10232929.9	a
		GG	6.78	6025595.9	b
	Il1b	GT	6.91	8128305.2	a
		TT	6.9	7943282.3	a
		GG	6.71	5128613.8	b
IgE	*Il1a*	TT	2.02	104.7	a
		TC	1.87	74.1	b
		CC	1.5	31.6	c
	Il1b	TT	1.92	83.2	a
		GT	1.77	58.9	b
		GG	1.59	38.9	c
	IL1	TTTT	2.02	104.7	a
		TCGT	1.91	81.3	ab
		TCTT	1.78	60.3	bc
		CCTT	1.67	46.8	bc
		CCGG	1.59	38.9	c
		CCGT	0.93	8.5	d
IgG	*H2Ab*	GG	3.88	7585.8	a
		AA	3.85	7079.5	a
		AG	3.73	5370.3	b
	H2Eb	GG	3.86	7244.4	a
		AA	3.82	6606.9	ab
		AG	3.71	5128.6	b
	H2	AAAA	3.9	7943.3	a
		AAGG	3.9	7943.3	a
		GGGG	3.88	7585.8	a
		GGAA	3.83	6760.8	ab
		GGAG	3.81	6456.5	ab
		AGGG	3.74	5495.4	ab
		AGAG	3.71	5128.6	ab
		AAAG	3.55	3548.1	ab
Macrophages	*H2Aa*	GG	5.86	724436.0	a
		AA	5.84	691831.0	ab
		AG	5.58	380189.4	b

Table 4.9. The results from post-hoc Tukey's tests demonstrating the specific genotypic associations to the log_{10} cell phenotypes of the sampled mice

Discussion

SNP Identification

H2-Locus

The genes of the MHC locus (class I and class II) are the most well studied immunological genes among mammalian species (Acevedo-Whitehouse and Cunningham, 2006). Numerous studies have identified genetic associations to specific polymorphisms in these genes to variation in immune function (reviewd in Sommer, 2005), often measured as resistance or susceptibility to specific infections. In house mice, several associations have been identified, such as the effect of *H2* haplotypes on IgA myeloma proteins (Lieberman and Humphrey, 1971). However, the genetic variances in the genes of the *H2*-locus are often poorly characterised, described by their allotype or haplotype rather than the specific genetic variant. In the three *H2*-locus genes sequenced in this study, 11 SNPs were previously reported among the common inbred strains of mice. However, none of these SNPs were reported to be functionally significant.

The number / proportion of non-synonymous SNPs in the *H2* loci compared to the non-*H2* loci is the most striking difference in considering the genetic diversity within these genes. 16/33 SNPs in exons were identified as being non-synonymous, 48%. For *H2*-Aa 25%, *H2*-Ab 57% and *H2*-Eb 71%. The high presence of non-synonymous SNPs supports the theory of balancing selection. In addition, non-synonymous SNPs in H-2 loci had high frequency among the sampled wild mice, up to 10/23 for *H2*-Ab and 11/23 for *H2*-Eb. The frequency of non-synonymous SNPs in *H2*-Aa was lower with a high of only 2/11. These findings suggest that there is positive selection for substitutions in the loci (Garrigan and Hedrick, 2003; Sommer, 2005; Piertney and Oliver, 2006). Similarly, to the proteins produced by the *H2*-locus toll-like receptor proteins (TLRs) must detect pathogens and trigger signalling cascades in response. However, the pattern recognition regions of these genes are much more conserved than genes in the *H2*-locus (Aderem and Ulevitch, 2000; Medzhitov and Janeway, 2000) and therefore it is surprising that among the three TLR genes sequenced in the current study 4/6 SNPs in the exon regions of TLR4 were non-synonymous and the

single SNP identified in *Tlr9* was also non-synonymous. However, when KASP genotyping was carried out at these loci the TLR4 locus was monomorphic, and the *Tlr9* SNP had an allele frequency of only 0.05. Therefore, from the current data it is difficult to make comparisons between the *H2*-loci and TLRs and it is difficult to determine whether the SNP loci in the *Tlr* genes are real without repeats which was beyond the limits of the current study.

Despite the relatively small scope of this study in comparison to human and domestic livestock studies, a tremendous amount of novel variation within the three loci of the *H2*-locus has been identified suggesting that further investigation of these loci in wild mice may yield even more variants. This rich source of novel variation and the evidence that some of these variants have functional significance to elements of immune functions prove that immunogenetic investigation into wild species, particularly house mice, may be a crucial tool in further understanding the role of the MHC complex in mammalian immune function within a natural, wild context.

Heterozygosity

Increasing heterozygosity within the three *H2*-loci studied did not significantly affect the number of infections the house mouse possessed. There were very few mice that were heterozygote at all three loci examined, which may suggest that this genotype is selected against. However, heterozygosity in at least one of the three loci is significantly associated with resistance to infections. These analyses were relatively simplistic in their approach and have not considered the effects of linked, non-genotyped, loci or resistance to a specific infection.

Non H2-Locus

As expected, the non-*H2* immune loci examined in the current study demonstrated a greatly reduced amount of genetic variation compared to their *H2*-locus counterparts. Many loci were either monomorphic, or the SNP investigated was rare, having a very low minor allele frequency. However, the lack of variation in these genes may actually be due to several potential study design flaws. First the number of mice sequenced

for SNP discovery was very low with only a few individuals from each sub-population represented, meaning that very little of the potential variation in these loci could be capture, in addition to this the sample size of genotyped mice was also small, especially considering the population sub-structuring that we now know is present among the mice in this sample. The Sanger sequencing that was performed on the sampled mice was also limited with only single reads carried out and for most samples no repeats were conducted to confirm the SNPs. Therefore, it is likely that some of the mono-morphic SNPs identified from the KASP genotyping were because of false positives reported from the Sanger sequencing, some SNPs were only present in a single mouse from sequencing.

Several non-*H2* loci displayed a considerable amount of polymorphism in the number of SNPs reported following sequencing, although again this may in some part be due to false positives reported from single read sequencing. However, in the exon regions of *Tnf*, *Tlr4*, *Il17a* and *Il13* at least two SNPs were present in multiple mice yet only the two SNPs from *Il17a* (20733797 and 20734091) were polymorphic when KASP genotyped. This may suggest that the monomorphism displayed in the other loci following KASP genotyping may be due to an error in the methodology, perhaps in the design of the KASP genotyping primer sequences but would require repeating to confirm this.

The ability of this study to identify novel genetic variants in the non-*H2* locus, several of which have significant associations to measures of immune function, demonstrates the importance of wild organisms for identifying important genetic variants. In addition, the number of reported SNPs in the non-*H2* loci suggest that these loci may be more variable than traditional laboratory-based studies suggest, and this variation may be a significant factor in the survival of wild organisms due to their effect on the hosts immune function. Particularly wild house mice present an excellent tool for further investigating the effects of these novel variants as significant SNPs identified from the wild mice can be induced into laboratory strains of mice through genetic engineering and their effect then measured in a controlled environment, including a controlled genetic background, to minimise confounding effects. However, it is

important to still consider that the wild environment and the context of an organism's immune system may be important for developing unique phenotype to which genetic associations can be sought.

Comparisons to Myo1a

Myo1a codes for myosin protein with no previously reported association to immune function and was therefore selected as a predicted neutral non-immune comparison. A total of 12 SNP loci were successfully sequenced from this gene, two in exons and 10 in introns, both of the exon SNPs were synonymous. However, when KASP genotyping was carried out on one of the exons SNPs and intron SNPs for this gene the exon SNP had a minor allele frequency of only 0.06 and therefore below the established threshold to carryout association analyses. In comparison the selected intron SNP of this gene (127710617 T/G) had a much higher minor allele frequency of 0.35.

Genetic Associations

A total of seven direct genetic associations were identified for both *H2* and non-*H2* loci (*H2*Eb, *H2*Ab, Il1a and Il1b) all of which are novel for house mice despite over 100 years of research in this species. Sex was a significant factor for IgE antibody titres only, with females having higher titres which is consistent with other studies of house mice (Astorquiza *et al.*, 1987), however, IgE titres were measured following antigen challenge with PHA in the previous study. For $CD8^+$ T cells age was the only factor that affected the scaled cell count. Site effects were significant in nine immune phenotypes, as discussed in a previous chapter the populations of house mice are highly sub-divided. This subdivision is a result of geographical isolation which therefore may lead to significant variations in the environments these mice live and in particular the infections they experience. The effect of this sub-division as a confounding factor on genotype (site by genotype) is also evident in several significant association, for B cells *H2*Ab and Myo1a, for IgE *H2*Aa and for IgA Myo1a.

No direct genetic associations to this locus and immune function were identified among the sampled mice, however there were significant site-by-genotype associations for B cells and IgA. However, it may be possible that these site-by-genotype effects are actually caused by linkage between *Myo1a* and the adjacent *Tac2* gene, which is known to be involved in pulmonary inflammation as well as its role in the central nervous system (NCBI Database, Mus musculus; Gene ID:21334)

Inferred Mode of Inheritance

H2Eb and *Il1b* significantly influence the scaled cell count of B cells, homozygous GG individuals have a significantly lower scaled cell count than either the alternative homozygotes or heterozygotes suggesting that there is a dominant / recessive mode of inheritance. For IgE the *Il1a* and *Il1b* loci each of the genotypes significantly differs from one another in an additive manner, with the heterozygote individuals having the intermediate value if IgE blood titres suggesting a potential codominant mode of inheritance. The combined IL1 locus is significantly associated with IgE titres (0.001). The combined effect of IL1 genotypes on IgE suggests that the presence of a T allele at the il1a locus has the strongest effect on antibody titres. However, as with the combined *H2* loci not all possible genotypes were present among the sampled mice and therefore3 the full effects each genotype cannot be determined. For IgG titre *H2*Ab and *H2*Eb were displayed similar patterns of genotypic effect. For both the heterozygotes exhibited significantly lower titres of antibodies compared to one of the homozygotes (GG) but overlapped with the other homozygote. The combined *H2* locus did not have a significant effect on IgG titres (0.18) and in addition no pattern could be distinguished among the different genotypes of the combined loci.

Conclusion

Wild house mouse populations are an untapped source of novel genetic variation that may allow for new insights to be drawn into the role of genetic variation and the immune system. Genetic diversity among wild house mice is considerably higher than among laboratory house mice as demonstrated by the high number of novel SNPs discovered in previously sequenced regions. The *H2*-locus is known to be particularly variable in house mice, however this study identified many more novel polymorphisms that are yet to be reported among laboratory house mice. In addition, considerable genetic variation was also identified in non-*H2*-locus genes, particularly within IL-1b which is relatively homogeneous among laboratory house mice.

The genetic associations identified in this study are also novel, in that they are the first to statistically associate phenotypic variation in these immunological traits to genetic variance, and in the methodology to do so. The specific associations identified warrant further investigation in laboratory studies to understand whether the is a causal relationship between these variants and phenotypic variance. The fact that these associations can be identified in this population of wild house mice also suggests that the study of immunogenetics in wild populations, particularly mice, can provide new insights into the effects of genetics and the environment on immune function. As humans experience both genetic diversity and environmental variation, studying house mice, our greatest tool in immunogenetic research, in the wild will give mouse research data greater relevance to human populations.

Chapter 5: General Discussion

Population Genetics of Wild Mice

The findings of this study clearly demonstrate that populations of wild house mice in southern Britain are highly subdivided. Even among sites that are in close geographic proximity to each other there is strong, significant, genetic subdivision. These findings are consistent with past work on commensal house mice, showing that mice are organised into small demes and rarely migrate beyond 25m within their lifetime (Lewontin and Dunn, 1960). Therefore, there is little geneflow between these demes. A surprising find of this study was the extent to which this gene flow is limited between adjacent farms with, for example, Barrow Mill and St Katherine's being located less than 200m metres apart but with mice at these sites having a significant F_{ST} value ($p<0.001$). Further to this several migrant mice were identified between these two sites, possessing genotypes distinct to their site of origin and not site of capture (supported by the STRUCTURE finding too), demonstrating that migration between sites can occur but these migrants make limited contributions to the host demes gene pool. The F_{ST} values reported in this study are substantial for a mammalian population with an average value of 0.461. In comparison to previous F_{ST} data on wild house mice the value from the current study is much higher, average value 0.172, although only three allozymes were used and the study was conducted in the USA.

Isolation-by-distance was not shown to be significant in this study despite the high F_{ST} values and other measures of population sub-division. This is because the geographic distances among site mice were sampled from are beyond the threshold of near-complete genetic isolation. This results in populations having similarly high genetic distances despite large variations in their geographic distances. Therefore, to study isolation-by-distance in more detail future studies should use sites that are located within a few kilometres of each other. It may also be possible to identify the threshold of genetic isolation by sampling and genotyping mice at set distance intervals; however, this is unlikely to be possible due to sporadic distribution of mouse populations. As discussed, the high F_{ST} values may also have been inflated due to the

presence of related mice from multiple generations being included in the sampling. In order to calculate more accurate estimates of genetic sub-division future studies should try to remove this confounding variable. This would depend on the ability to identify individuals and their offspring which is not easy to achieve in a wild population. One way to reduce this problem could be to only sample each population once, weight or age could then be used as a criterion to identify individuals that are not yet sexually mature, removing the risk of multiple generations being sampled. However, this method would severely reduce the sample size of the study. Another possibility could involve an arena study where an artificial environment simulating wild condition is seeded with wild captured mice. However, this would be expensive and on a very small scale compared to the current study.

Mice from the London underground (LU) were also sampled in this study. Two important conclusions can be drawn from the population genetic analyses involving these mice. Firstly, despite the mice being sampled from different stations there is a greater homology among the LU mice than expected, demonstrated by the strong clustering of these mice and their similar branch lengths compared to other sub-populations in a bootstrapped nearest-neighbour joining tree analysis. Secondly the LU mice have far higher gene flow compared to mice in the Bristol sub-populations (BR) and mice from Skokholm island (SK), despite the similar geographical distances between these locations. The pairwise F_{ST} values between LU and BR sub-populations were much smaller in comparison to the BR-SK comparisons, and in some cases the F_{ST} values between the LU and BR sub-populations were smaller than between some of the BR sub-populations themselves. Therefore, it is likely that the London underground transport system and its connection to the national rail network provides migratory paths for these mice. However, it is unclear as to what direction this gene flow occurs. Further studies of these mice involving following unique genetic markers from one sub-population to another would help to reveal the rate of gene flow and the direction. Mice from the SK sub-population were the most genetically distinct of the wild mice sampled, having the highest pairwise F_{ST} values when compared to all other sub-populations and in the nearest-neighbour joining tree and STRUCTURE analyses these mice clustered the most strongly together, demonstrating

their genetic isolation to the other mice. It is not surprising that this sub-population is the most genetically distinct as it is geographically distant and with a body of water as a migration obstacle. However, what is interesting about this sub-population is the amount of homogeneity within it. In comparison to the other house mice populations there is a much lower amount of heterozygosity on the island and the relatively small length of the branches on the nearest-neighbour joining tree demonstrate that there is a high degree of inbreeding on the island despite mice being sampled from different locations on the island. These findings correspond with data for feral wild house mice and their increased propensity to migrate greater distances, abandon deme structures and therefore have a greater amount of geneflow within the population (Berry, 1964; Reimer and Petras, 1967; Jones *et al.*, 1995). The mice of Skokholm provide a unique opportunity to compare and contrast laboratory, and commensal-wild house mice to and model the differences between urban and rural living in a model organism. This may be of particular importance in immunology with the concepts of the hygiene hypothesis and the contrasts between allergy and autoimmunity in urban and rural populations.

The most likely contributor to the genetic isolation and a barrier to geneflow between sub-populations is the loss of hedgerows, ditches and thickets that used to connect farms and settlements and provided migratory routes for small organisms including house mice (Sellers *et al.*, 2018). Over the past 50 years there has been a continuous reduction in these forms of cover (Kotzageorgis and Mason, 1997). In addition, changes in farming practices and the increase in pet ownership and therefore predation on small mammals likely means that migration is becoming harder to achieve (Brown *et al.*, 2004). There is very limited data comparing the genetic diversity of wild house mice over time and those that have been carried out used a small number of genetic variants to determine diversity. Therefore, a longitudinal study of wild house mice in Britain may be an important addition to our ecological understanding of these animals and how changes in their ecosystems have affected their genetic diversity and population sizes. Specifically, adjacent farms with and without hedgerows can be studied to compare the effect of hedgerows on gene flow.

Mouse populations from Skokholm Island and the London Underground present novel opportunities to examine house mice and compare commensal and feral lifestyles. To test the theory that there is greater gene flow in a population the seeding of mice on Skokholm Island, carrying unique genetic markers, would allow for the calculation of migration and gene flow in such populations (Anderson *et al.*, 1964; Baker, 1981).

The London Underground present a novel opportunity to study wild house mice living under high anthropogenic influence and with unique migratory routes. An interesting future study could investigate how geneflow among the mice varies between stations and tube lines. The genotypic differences could then be mapped and overlaid with the physical distances to see if they match.

Previous population genetic studies in wild house mice have been more limited in their genotyping capacity, mainly using microsatellite or allozyme data, and therefore gave a narrow view of the geneflow. The current study has analysed genotypic data across the genome of the house mouse to give a more accurate account of the population genetics of a wild mammal in its ecological context.

The Heritability of Immune Function in Wild House Mice

The heritability of immunological traits is highly variable (Aguirre-Gamboa *et al.*, 2016). Changes in species or methodology can give drastically different results. In the current study I show that several measures of immune function do have significant heritabilities. This is an important result when considering the population tested was wild and therefore experiencing greater environmental variation than laboratory studies, and the genetic relatedness was unknown. In particular the pathogenic challenges these organisms face is highly variable and likely contributes significantly to the observed variation in immune function among mice. Several studies have recently reported the significance of studying immune function in a natural context (Abolins *et al.*, 2011; Boysen *et al.*, 2011), as novel phenotypes identified in some wild populations have not been found in their domesticated or laboratory counterparts.

Despite the importance of this current study in identifying the heritability of these immune phenotypes in a wild population these results should be taken with caution as there may be confounding site effect factors that could not be considered. When mice from a single sub-population were examined the heritabilities were substantially different. However, three phenotypes, IgG, IgE and IL12p70 were significantly heritable in both analyses with substantial heritability values. In order to confirm these heritabilities in wild mice a longitudinal study of wild house mice should be conducted, in which the actual familial relatedness of individuals can be identified, while still experiencing a wild environment. In addition, this longitudinal study should be carried out in a single large population, for example they Hyatt's Wood farm population, as this would remove the potentially confounding site effects. A longitudinal study of a single population would allow for a more accurate calculation of heritability in wild house mice as we would be more able to determine the relatedness of individuals through a pedigree or twin study. In order to determine relatedness mice would need to have physical markers to identify them the breeding of the mice monitored over time. A large sample size would give greater statistical power and the use of mice at the same age and from the same sample site would remove the need to use covariate estimates.

Immune Gene Genetic Diversity and Genetic Associations

This study is the first of its kind to look at immune gene diversity in wild house mice and builds on previous research carried out on wild voles in Britain (Turner, 2010; Turner *et al.*, 2011). Many of the SNPs identified in these genes are novel, having not previously been identified in laboratory bred house mice despite over 100 years of intensive study (Castle and Allen, 1903; Phifer-Rixey and Nachman, 2015). In particular the major histocompatibility locus has been studied in detail since the 1970's (Klein and Bailey, 1971), yet this study has revealed novel SNPs that have not previously been characterised in this species. Not only were novel SNPs identified but these SNPs had significant associations to variation in immune function. However, this association does not mean these SNPs are necessarily causal but could be in linkage to the causal SNPs. In order to determine the direct effect of these SNPs in isolation from other genetic variants, reverse genetic studies in house mice could be carried out. This would involve inducing the specific SNP into a mouse with a known genetic background and then comparing the immune phenotypes of these mice to others with the same background but no mutation. Regardless of whether these SNPs are causal or not, the study demonstrates the obvious reductionist limitations of laboratory-based studies in their inability to identify important genetic variants from a limited gene pool of inbred mice.

As expected in comparison to non-*H2* locus genes, the *H2*-loci had a far greater number of SNPs, and in addition, a greater number, and proportion, of these SNPS were non-synonymous. The sequencing SNP data suggests that the *H2*-locus genes are experiencing higher selective pressures for substitutions than in non-*H2*-locus genes, this is in order to maximise the number of different pathogenic antigens a host can respond to.

Unfortunately, a potential limitation with this study was that all of the data at the *H2*-locus was for class II receptors and none for the class I receptors. Two *H2*-locus class I genes (*H2*D1 and *H2*K1) were selected as candidate genes but failed to successfully

amplify by PCR and therefore could not be sequenced or genotyped. The cause of the failure to amplify is unknown, however it is possible that high sequence diversity in the regions to which the PCR primers were designed meant that the primers could not adequately anneal and therefore amplification was not achieved. Future studies may wish to redesign these primers in order to reliably amplify these genes for sequencing and genotyping to allow for comparisons to the *H2* class II genes. In addition to having high genetic diversity it was also revealed that mice that are heterozygous in at least two of the genotyped *H2*-loci have significantly reduced chance of becoming infected by the 7 assayed pathogens. This is in keeping with previous studies that suggest heterozygosity in the MHC is protective against pathogenic infections (Doherty and Zinkernagel, 1975; Nei and Hughes, 1991).

Despite the ongoing focus of immunogenetic and ecoimmunological studies on genes in the MHC, this study has revealed that not only do wild house mice possess considerable amounts of genetic variation in non-MHC loci, but that novel genetic associations are also waiting in obscurity in these wild populations. The undiscovered phenotypes may lead to new associations and disease mechanisms being discovered, allowing for new laboratory mouse models to be developed and ultimately may contribute to our understanding of human immune function in a "wild" context. In order to better understand the functional effects of these novel mutations transgenic lines of mice could be bred, carrying these specific polymorphisms. The resultant effect on immune function can then be studied under controlled conditions to determine to what extent the polymorphism affects specific phenotypes.

A limitation of this study was in the SNP discovery carried out. A very small sample of mice were selected for sequencing. While these mice were selected to be representative of the sub-populations sampled, for many sub-populations only a single mouse was used, meaning that unique genetic variants in these sub-population were likely missed. In addition, only single read Sanger sequencing was conducted, which allowed for the identification of a large number of novel SNPs, but also has a low reliability. Many of the SNPs discovered within the candidate genes had low representation among the sample mice sequenced. Therefore, apart from the SNPs

confirmed by the subsequent KASP genotyping, it is difficult to determine whether the SNPs identified are real or false positives due to sequencing errors. In order to rectify this, next generation sequencing techniques should be used on a much larger number of mice, drastically improving the accuracy and reliability by reducing the rate of false negatives and increasing the chances of identifying novel SNP substitutions. Ultimately GWAS studies in wild house mice would identify further genetic associations in immune function but require a larger population sample, with little or no relatedness between individuals and a large array of genetic markers across the genome and therefore have primarily been carried out in humans and livestock (Rivera and Tessarollo, 2008; Thompson-Crispi *et al.*, 2014; Uzzaman *et al.*, 2014; Brodin *et al.*, 2015). This is all possible in the house mouse, owing to its long career in biological research, which means we now have the resources to collect detailed phenotypic and genetic data from the wild. The variation found in this small-scale study demonstrates that wild house mice harbour important genetic variation that has so far not been possible to study in laboratory-bred mice, and better reflect humans as a model organism. In order to identify more genetic variants in wild house mice, in both immune and non-immune related genes a largescale sequencing project could be carried out. This would provide novel genetic markers that could then be used in a future GWAS of wild house mice, which will have significantly more power to identify genetic associations.

Finally, this study only examined cell counts and antibody titres as measures of immune function. While these measures represent humoral and cellular immune function, they are overly simplistic in the context of immune function overall. For a more in-depth perspective of genetic associations serum cytokine concentrations as well as cell population proportions could be examined in a future study.

Conclusion

This study presents evidence that wild house mice present a unique opportunity to study ecoimmunology and immunogenetics in realistic “natural” setting. Generations of intensive work as a laboratory model organism can now be successfully applied to wild house and allow us to close the gap in our knowledge off immunogenetics in a wild setting. Not only did this sample of wild house mice harbour considerable genetic variation overall but also in functional genes of their immune system, some of which have significant effects on measures of their immune function. In addition, this study demonstrated that the heritability of phenotypes can be successfully calculated from a wild population without detailed relatedness or pedigree data. These findings suggest that studies of wild house mice can have applications to population genetics, epidemiology, immunology, ecoimmunology, evolution and many more areas.

Bibliography

Abolins, S., King, E. C., Lazarou, L., Weldon, L., Hughes, L., Drescher, P., Raynes, J. G., Hafalla, J. C. R., Viney, M. E. and Rile, E. M. (2017) 'The comparative immunology of wild and laboratory mice, *Mus musculus domesticus*', *Nature Communications*. Nature Publishing Group, 8(May), pp. 1–13. doi: 10.1038/ncomms14811.

Abolins, S., Lazarou, L., Weldon, L., Hughes, L., King, E. C., Drescher, P., Pocock, M. J. O., Hafalla, J. C. R., Riley, E. M. and Viney, M. (2018) 'The ecology of immune state in a wild mammal, *Mus musculus domesticus*', *PLoS Biology*, 16(4). doi: 10.1371/journal.pbio.2003538.

Abolins, S. R., Pocock, M. J. O., Hafalla, J. C. R., Riley, E. M. and Viney, M. E. (2011) 'Measures of immune function of wild mice, *Mus musculus*.', *Molecular ecology*, 20(5), pp. 881–892. doi: 10.1111/j.1365-294X.2010.04910.x.

Acevedo-Whitehouse, K. and Cunningham, A. A. (2006) 'Is MHC enough for understanding wildlife immunogenetics?', *Trends in Ecology and Evolution*, 21(8), pp. 433–438. doi: 10.1016/j.tree.2006.05.010.

Aderem, A. and Ulevitch, R. J. (2000) 'Toll-like receptors in the induction of the innate immune response', *Nature*, 406(6797), pp. 782–787. doi: 10.1038/35021228.

Aguirre-Gamboa, R. *et al.* (2016) 'Differential effects of environmental and genetic factors on T and B cell immune traits', *Cell Reports*. The Authors, 17(9), pp. 2474–2487. doi: 10.1016/j.celrep.2016.10.053.

Alexander, E. S. *et al.* (2014) 'Twin and family studies reveal strong environmental and weaker genetic cues explaining heritability of eosinophilic esophagitis', *Journal of allergy and clinical immunology*, 135(4), pp. 1084-1092.e1. doi: 10.1038/jid.2014.371.

Alves, J. M., Carneiro, M., Afonso, S., Lopes, S., Garreau, H., Boucher, S., Allain, D., Queney, G., Esteves, P. J., Bolet, G. and Ferrand, N. (2015) 'Levels and patterns of genetic diversity and population structure in domestic rabbits', *Plos One*, 10(12), p. e0144687. doi: 10.1371/journal.pone.0144687.

Amos, W., Driscoll, E. and Hoffman, J. I. (2011) 'Candidate genes versus genome-wide associations: Which are better for detecting genetic susceptibility to infectious disease?', *Proceedings of the Royal Society B: Biological Sciences*, 278(1709), pp.

1183–1188. doi: 10.1098/rspb.2010.1920.

Anderson, P. K. (1964) 'Lethal alleles in *Mus musculus*: local distribution and evidence for isolation of demes.', *Science (New York, N.Y.)*, 145, pp. 177–178. doi: 10.1126/science.145.3628.177.

Anderson, P. K., Dunn, L. C. and Beasley, A. B. (1964) 'Introduction of a lethal allele into a feral house mouse population', *American Society of Naturalists*, 98(898), pp. 57–64.

Ardia, D. R. and Rice, E. B. (2006) 'Variation in heritability of immune function in the tree swallow', *Evolutionary Ecology*, 20(5), pp. 491–500. doi: 10.1007/s10682-006-0016-x.

Astorquiza, M. I., Cisternas, C. and Leal, X. (1987) 'Sex-dependent differences in the IgE response modulated by phytohemagglutinin', *Immunology Letters*, 16(1), pp. 27–30. doi: 10.1016/0165-2478(87)90056-3.

Baker, E. M. (1981) 'Gene flow in house mice : introduction of a new allele into free-living populations', *Evolution*, 35(2), pp. 243–258.

Barrett, J. C., Fry, B., Maller, J. and Daly, M. J. (2005) 'Haploview : analysis and visualization of LD and haplotype maps', *Bioinformatics*, 21(2), pp. 263–265. doi: 10.1093/bioinformatics/bth457.

Bennett, D., Bruck, R., Dunn, L. C., Klyde, B., Shutsky, F. and Smith, L. . (1967) 'Persistence of an introduced lethal in a feral house mouse population', *American Naturalist*, 101, pp. 538–539.

Bergman, I., Basse, P. H., Barmada, M. A., Griffin, J. A. and Cheung, N.-K. V. (2000) 'Comparison of in vitro antibody-targeted cytotoxicity using mouse, rat and human effectors', *Cancer Immunology, Immunotherapy*, 49(4–5), pp. 259–266. doi: 10.1007/s002620000120.

Berry, R. J. (1964) 'The evolution of an island population of the house mouse', *Evolution*, 18, pp. 468–483.

Berry, R. J. (1967) 'Genetical changes in ,mice and men', *Eugenics Review*, 52, pp. 78–96.

Berry, R. J. and Jakobson, M. E. (1974) 'Vagility in an island population of the house mouse', *Proceedings of the Zoological Society of London*, 173(3), pp. 341–354. doi: 10.1111/j.1469-7998.1974.tb04119.x.

Bonhomme, F., Rivals, E., Orth, A., Grant, G. R., Jeffreys, A. J. and Bois, P. R. J. (2007) 'Species-wide distribution of highly polymorphic minisatellite markers suggests past and present genetic exchanges among house mouse subspecies', *Genome Biology*, 8(5). doi: 10.1186/gb-2007-8-5-r80.

Bonhomme, F., Searle, J. B. and Pialek, J. (2012) 'House mouse phylogeography', in Macholan, M., Baird, S. J. E., and Munclinger, P. (eds) *Evolution of the House Mouse*. Cambridge: Cambridge University Press, pp. 278–296. doi: 10.1017/CBO9781139044547.013.

Borghans, J. A. M., Beltman, J. B. and De Boer, R. J. (2004) 'MHC polymorphism under host-pathogen coevolution', *Immunogenetics*, 55(11), pp. 732–739. doi: 10.1007/s00251-003-0630-5.

Boursot, P., Din, W., Anand, R., Darvich, D., Dod, B., Von Deimling, F., Talwar, G. P. and Bonhomme, F. (1996) 'Origin and radiation phylogeny of the house mouse: mitochondrial', *Journal of Evolutionary Biology*, 9, pp. 391–415. doi: 10.1046/j.1420-9101.1996.9040391.x.

Boysen, P., Eide, D. a G. M. and Storset, A. K. (2011) 'Natural killer cells infFree-living *Mus musculus* have a primed phenotype.', *Molecular ecology*, 20(23), pp. 5103–5110. doi: 10.1111/j.1365-294X.2011.05269.x.

Brodin, P. and Davis, M. M. (2017) 'Human immune system variation', *Nature Reviews Immunology*, 17(1), pp. 21–29. doi: 10.1038/nri.2016.125.

Brodin, P., Jojic, V., Gao, T., Bhattacharya, S., Lopez Angel, C. J., Furman, D., Shen-Orr, S., Dekker, C. L., Swan, G. E., Butte, A. J., Maecker, H. T. and Davis, M. M. (2015) 'Variation in the human immune system is largely driven by non- heritable influences', 160(0), pp. 37–47. doi: 10.1109/TMI.2012.2196707.Separate.

Brown, E. a, Pilkington, J. G., Nussey, D. H., Watt, K. a, Hayward, a D., Tucker, R., Graham, a L., Paterson, S., Beraldi, D., Pemberton, J. M. and Slate, J. (2013) 'Detecting genes for variation in parasite burden and immunological traits in a wild population: testing the candidate gene approach.', *Molecular ecology*, 22(3), pp. 757–73. doi: 10.1111/j.1365-294X.2012.05757.x.

Brown, P. R. and Singleton, G. R. (2002) 'Impacts of house mice on crops in Australia—costs and damage', in *In: Clark, L. (Ed.), Human– Wildlife Conflict: Economic Considerations.*, pp. 48–58.

Brown, R., Davies, M. J., Singleton, G. R. and Croft, J. D. (2004) 'Can farm-management practices reduce the impct of house mouse populations on crops in an irrigated farming system', *Wildlife Research*, 31(6), pp. 597–604.

Cassetta, L., Cassol, E. and Poli, G. (2011) 'Macrophage polarization in health and disease', *The Scientific World JOURNAL*, 11, pp. 2391–2402. doi: 10.1100/2011/213962.

Van De Casteele, T., Galbusera, P. and Matthysen, E. (2001) 'A comparison of microsatellite-based pairwise relatedness estimators', *Molecular Ecology*, 10(6), pp. 1539–1549. doi: 10.1046/j.1365-294X.2001.01288.x.

Castillo-Fernandez, J. E., Spector, T. D. and Bell, J. T. (2014) 'Epigenetics of discordant monozygotic twins: Implications for disease', *Genome Medicine*, 6(7), pp. 1–16. doi: 10.1186/s13073-014-0060-z.

Castle, A. W. E. and Allen, G. M. (1903) 'The heredity of albinism', *Proceedings of the American Academy of Arts and Sciences*, 38(21), pp. 603–622.

Clapperton, M., Glass, E. J. and Bishop, S. C. (2008) 'Pig peripheral blood mononuclear leucocyte subsets are heritable and genetically correlated with performance', *Animal*, 2(11), pp. 1575–1584. doi: 10.1017/S1751731108002929.

Coltman, D. W. (2005) 'Testing marker-based estimates of heritability in the wild', *Molecular Ecology*, 14(8), pp. 2593–2599. doi: 10.1111/j.1365-294X.2005.02600.x.

Coltman, D. W., Wilson, K., Pilkington, J. G., Stear, M. J. and Pemberton, J. M. (2001) 'A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in a naturally-parasitized population of Soay sheep', *Parasitology*, 122, pp. 571–582.

Cotter, S. and Wilson, K. (2002) 'Heritability of immune function in the caterpillar *Spodoptera littoralis*', *Heredity*, 88, pp. 229–234. doi: 10.1038/sj/hdy/6800031.

de Craen, A. J. M., Posthuma, D., Remarque, E. J., van den Biggelaar, A. H. J., Westendorp, R. G. J. and Boomsma, D. I. (2005) 'Heritability estimates of innate immunity: An extended twin study', *Genes and Immunity*, 6(2), pp. 167–170. doi: 10.1038/sj.gene.6364162.

Crowcroft, P. and Rowe, F. P. (1963) 'Social organisation and territorial behaviour in the wild house mouse (*Mus musculus L.*)', *Proceedings of the Zoological Society of London*, 140(3), pp. 517–531. doi: 10.1111/j.1469-7998.1963.tb01871.x.

Cruzan, M. B. (1998) 'Genetic markers in plant evolutionary ecology', *Ecology*, pp. 400–412. doi: 10.1890/0012-9658(1998)079[0400:GMIPEE]2.0.CO;2.

Daley, J. G. (1992) 'Population reductions and genetic variability in black-tailed prairie dogs', *Journal of Wildlife Management*, 56(2), pp. 212–220.

Dallas, J. F., Dod, B., Boursot, P., Prager, E. M. and Bonhomme, F. (1995) 'Population subdivision and geneflow in Danish house mice', *Molecular Ecology*, 4(3), pp. 311–320.

Damania, B. and Dittmer, D. P. (2014) 'What lies within: coinfections and immunity', *Cell Host & Microbe*, 16(2), pp. 145–147. doi: 10.1016/j.chom.2014.07.014.

Decanini, L. I., Collins, A. M. and Evans, J. D. (2007) 'Variation and heritability in immune gene expression by diseased honeybees', *Journal of Heredity*, 98(3), pp. 195–201. doi: 10.1093/jhered/esm008.

DeFries, J. C. and McClearn, G. E. (1972) 'Behavioral genetics and the fine structure of mouse populations- a study in microevolution', *Evolutionary Biology*, 5, pp. 279–291.

Dietze, K. K., Dittmer, U., Koudaimi, D. K., Schimmer, S., Reitz, M., Breloer, M. and Hartmann, W. (2016) 'Filariae-Retrovirus co-infection in mice is associated with suppressed virus-specific IgG immune response and higher viral loads', *PLoS Neglected Tropical Diseases*, 10(12), pp. 1–17. doi: 10.1371/journal.pntd.0005170.

Din, W., Boursot, P., Anand, R., Darvich, D., Dod, B., Von Deimling, F., Talwar, G. P. and Bonhomme, F. (1996) 'Origin and radiation phylogeny of the house mouse: mitochondrial', *Journal of Evolutionary Biology*, 9, pp. 391–415. doi: 10.1046/j.1420-9101.1996.9040391.x.

Doherty, P. C. and Zinkernagel, R. M. (1975) 'Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex', *Nature*, 256(5512), pp. 50–52. doi: 10.1038/256050a0.

Earl, D. A. and Bridgett, M. (2012) 'STRUCTURE HARVESTER : A website and program for visualizing STRUCTURE output and implementing the Evanno method', *Conservation Genetic Resources*, 4, pp. 359–361. doi: 10.1007/s12686-011-9548-7.

Evanno, G., Regnaut, S. and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study', *Molecular Ecology*, 14(8), pp. 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x.

Excoffier, L. and Lischer, H. E. L. (2010) 'Arlequin suite ver 3 . 5 : a new series of programs to perform population genetics analyses under Linux and Windows', *Molecular Ecology*, 10, pp. 564–567. doi: 10.1111/j.1755-0998.2010.02847.x.

Fan, B., Wang, Z., Li, Y., Zhao, X., Liu, B., Zhao, S., Yu, M., Li, M., Chen, S., Xiong, T. and Li, K. (2002) 'Genetic variation analysis within and among chinese indigenous swine populations using microsatellite markers', *Animal Genetics*, 33, pp. 422–427.

Ferreira, M. A. R. (2018) 'Ten years of genome-wide association studies of immune-related diseases', *Clinical and Translational Immunology*, 7(6), pp. 1–3. doi: 10.1002/cti2.1022.

Ferris, S. D., Sage, R. D. and Wilson, A. C. (1982) 'Single female origin of inbred mice', *Nature*, 295, pp. 163–165.

Festa-Bianchet, M. (1989) 'Individual differences, parasites and the costs or reproduction for bighorn ewes (*Ovis canadensis*)', *The Journal of animal ecology*, 58, pp. 785–795.

Flori, L., Gao, Y., Laloë, D., Lemonnier, G., Leplat, J. J., Teillaud, A., Cossalter, A. M., Laffitte, J., Pinton, P., de Vaureix, C., Bouffaud, M., Mercat, M. J., Lefèvre, F., Oswald, I. P., Bidanel, J. P. and Rogel-Gaillard, C. (2011) 'Immunity traits in pigs: Substantial genetic variation and limited covariation', *PLoS ONE*, 6(7). doi: 10.1371/journal.pone.0022717.

Foll, M. and Gaggiotti, O. (2008) 'A Genome-scan method to identify selectedl loci appropriate for both dominant and codominant markers: A Bayesian perspective', *Genetic Society of America*, 993(October), pp. 977–993. doi: 10.1534/genetics.108.092221.

Fournier-Level, A., Korte, A., Cooper, M. D., Nordborg, M., Schmitt, J. and Wilczek, a. M. (2011) 'A map of local adaptation in *Arabidopsis thaliana*', *Science*, 334(6052), pp. 86–89. doi: 10.1126/science.1209271.

Fraga, M. F. *et al.* (2005) 'Epigenetic differences arise during the lifetime of monozygotic twins', *PNAS*, 102(30), pp. 10604–10609. doi: 10.1073/pnas.0500398102.

Freeman, Jr., R. B. (2009) 'The "Indirect" effects of Cytomegalovirus infection', *American Journal of Transplantation*, 9(11), pp. 2453–2458. doi: 10.1111/j.1600-6143.2009.02824.x.

French, S. S., DeNardo, D. F. and Moore, M. C. (2007) 'Trade-offs between the reproductive and immune systems: facultative responses to resources or obligate responses to reproduction?', *The American Naturalist*, 170(1), pp. 79–89. doi: 10.1086/518569.

Gabriel, S. I., Mathias, M. D. L. and Searle, J. B. (2013) 'Genetic structure of house mouse (*Mus musculus* Linnaeus 1758) populations in the Atlantic archipelago of the Azores: colonization and dispersal', *Biological Journal of the Linnean Society*, 108(4), pp. 929–940. doi: 10.1111/bij.12015.

Gabriel, S. I., Mathias, M. L. and Searle, J. B. (2015) 'Of mice and the "age of discovery": The complex history of colonization of the Azorean archipelago by the house mouse (*Mus musculus*) as revealed by mitochondrial DNA variation', *Journal of Evolutionary Biology*, 28(1), pp. 130–145. doi: 10.1111/jeb.12550.

Garrigan, D. and Hedrick, P. W. (2003) 'Perspective: detecting adaptive molecular polymorphism: lessons from the MHC', *Evolution*, 57(8), pp. 1707–1722. doi: 10.1111/j.0014-3820.2003.tb00580.x.

Gauly, M. and Erhardt, G. (2001) 'Genetic resistance to gastrointestinal nematode parasites in Rhön sheep following natural infection', *Veterinary Parasitology*, 102(3), pp. 253–259. doi: 10.1016/S0304-4017(01)00530-1.

Gay, L., Siol, M. and Ronfort, J. (2013) 'Pedigree-free estimates of heritability in the wild: promising prospects for selfing populations', *PLoS ONE*, 8(6). doi: 10.1371/journal.pone.0066983.

Glaubitz, J. C. (2004) 'CONVERT : A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages', *Molecular Ecology*, 4, pp. 309–310. doi: 10.1111/j.1471-8286.2004.00597.x.

Goodnight, K. F. and Queller, D. C. (1999) 'Computer software for performing likelihood tests of pedigree relationship using genetic markers', *Molecular Ecology*, 8, pp. 1231–1234. doi: 10.1007/s10336-008-0336-4.

Gordon, J. W. and Ruddle, F. H. (1981) 'Integration and stable germ line transmission of genes injected into mouse pronuclei', *Science*, 214(4526), pp. 1244–1246.

Graham, A. L., Hayward, A. D., Watt, K. a, Pilkington, J. G., Pemberton, J. M. and Nussey, D. H. (2010) 'Fitness correlates of heritable variation in antibody responsiveness in a wild mammal', *Science*, 330(6004), pp. 662–665. doi:

10.1126/science.1194878.
Gregersen, PK Olsson, L. (2010) 'Recent advances in the genetics of autoimmune disease', *Annu Rev Immunol*, 27, pp. 363–391. doi: 10.1146/annurev.immunol.021908.132653.Recent.
Grjibovski, A. M., Olsen, A. O., Magnus, P. and Harris, J. R. (2007) 'Psoriasis in Norwegian twins: contribution of genetic and environmental effects', *Journal of the European Academy of Dermatology and Venereology*, 21(10), pp. 1337–1343. doi: 10.1111/j.1468-3083.2007.02268.x.
Guénet, J. L. and Bonhomme, F. (2003) 'Wild mice: an ever-increasing contribution to a popular mammalian model', *Trends in Genetics*, 19(1), pp. 24–31. doi: 10.1016/S0168-9525(02)00007-0.
Gulland, F. M. D., Albon, S. D., Pemberton, J. M., Moorcroft, P. R. and Clutton-Brock, T. H. (1993) 'Parasite-associated polymorphism in a cyclic ungulate population', *Proceedings of the Royal Society B: Biological Sciences*, 254(1339), pp. 7–13. doi: 10.1098/rspb.1993.0119.
Hadley, C. (2006) 'Food allergies on the rise? Determining the prevalence of food allergies, and how quickly it is increasing, is the first step in tackling the problem', *EMBO Reports*, 7(11), pp. 1080–1083. doi: 10.1038/sj.embor.7400846.
Hardouin, E. A., Orth, A., Teschke, M., Darvish, J., Tautz, D. and Bonhomme, F. (2015) 'Eurasian house mouse (*Mus musculus L*.) differentiation at microsatellite loci identifies the Iranian plateau as a phylogeographic hotspot', *BMC Evolutionary Biology*, 15(1), pp. 1–12. doi: 10.1186/s12862-015-0306-4.
Haskell, M. J., Simm, G. and Turner, S. P. (2014) 'Genetic selection for temperament traits in dairy and beef cattle', *Frontiers in Genetics*, 5(OCT), pp. 1–18. doi: 10.3389/fgene.2014.00368.
Hayward, A. D. (2013) 'Causes and consequences of intra- and inter-host heterogeneity in defence against nematodes', *Parasite Immunology*, 35(11), pp. 362–373. doi: 10.1111/pim.12054.
Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., Maes, G. E., Diopere, E., Carvalho, G. R. and Nielsen, E. E. (2011) 'Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges', *Molecular Ecology Resources*, 11(SUPPL. 1), pp. 123–

136. doi: 10.1111/j.1755-0998.2010.02943.x.
Höhler, T., Reuss, E., Adams, P., Bartsch, B., Weigmann, B., Wörns, M., Galle, P. R., Victor, A. and Neurath, M. F. (2005) 'A genetic basis for IFN-gamma production and T-bet expression in humans.', *Journal of immunology (Baltimore, Md. : 1950)*, 175(8), pp. 5457–62. doi: 175/8/5457 [pii].
Hughes, A. L. (1991) 'MHC polymorphism and the design of captive breeding programs', *Conservation Biology*, 5(2), pp. 249–251. doi: 10.1111/j.1523-1739.1991.tb00130.x.
Hughes, A. L. and Hughes, M. K. (1995) 'Natural selection on the peptide-binding regions of major histocompatibility complex molecules', *Imununogenetics*, 42(4), pp. 233–243.
Hughes, C. (1998) 'Integrating molecular techniques with field methods in studies of social behavior: a revolution results', *Ecology*, pp. 383–399. doi: 10.1890/0012-9658(1998)079[0383:IMTWFM]2.0.CO;2.
Iacolina, L., Scandura, M., Goedbloed, D. J., Alexandri, P., Crooijmans, R. P. M. A., Larson, G., Archibald, A., Apollonio, M., Schook, L. B., Groenen, M. A. M. and Megens, H.-J. (2016) 'Genomic diversity and differentiation of a managed island wild boar population', *Heredity*. Nature Publishing Group, 116(1), pp. 60–67. doi: 10.1038/hdy.2015.70.
Jackson, J. A. (2015) 'Immunology in wild nonmodel rodents: An ecological context for studies of health and disease', *Parasite Immunology*, 37(5), pp. 220–232. doi: 10.1111/pim.12180.
Jakobsson, M. and Rosenberg, N. A. (2007) 'CLUMPP : A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure', *Bioinformatics*, 23(14), pp. 1801–1806. doi: 10.1093/bioinformatics/btm233.
Janss, L. L. G. and Bolder, N. M. (2000) 'Heritabilities of and genetic relationships between salmonella resistance traits in broilers', *Journal of animal science*, 78(9).
Jensen, J. L., Bohonak, A. J. and Kelley, S. T. (2005) 'Isolation by distance , web service', *BMC Genetics*, 6, pp. 1–6. doi: 10.1186/1471-2156-6-13.
Jones, A. G. and Ardren, W. R. (2003) 'Methods of parentage analysis in natural populations', *Molecular Ecology*, 12(10), pp. 2511–2523. doi: 10.1046/j.1365-

294X.2003.01928.x.

Jones, C., L, N., Jones, J., Tegelstrom, H., Triggs, G. and Berry, R. J. (1995) 'Differential male genetic success determines gene flow in an experimentally manipulated mouse population', *Proceedngs of the Royal Society, London*, 260, pp. 251–256.

Jones, E. P., Jensen, J. K., Magnussen, E., Gregersen, N., Hansen, H. S. and Searle, J. B. (2011) 'A molecular characterization of the charismatic Faroe house mouse', *Biological Journal of the Linnean Society*, 102(3), pp. 471–482. doi: 10.1111/j.1095-8312.2010.01597.x.

Kalinowski, S. T. (2011) 'The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure.', *Heredity*. Nature Publishing Group, 106(4), pp. 625–32. doi: 10.1038/hdy.2010.95.

Kinnard, T. B. and Westneat, D. F. (2009) 'Phenotypic and genetic variance of house sparrows (*Passer domesticus*) early in development', *The Auk*, 126(4), pp. 884–895. doi: 10.1525/auk.2009.08119.

Klaper, R., Ritland, K., Mousseau, T. a and Hunter, M. D. (2001) 'Heritability of phenolics in *Quercus laevis* inferred using molecular markers.', *The Journal of heredity*, 92, pp. 421–426. doi: 10.1093/jhered/92.5.421.

Klein, J. A. N. and Bailey, D. W. (1971) 'in natural populations of the house mouse 1', *Group*, (June).

Kochi, Y. (2016) 'Genetics of autoimmune diseases: perspectives from genome-wide association studies', *International Immunology*, 28(4), pp. 155–161. doi: 10.1093/intimm/dxw002.

Kopelman, N. M., Mayzel, J., Jakobsson, M. and Rosenberg, N. A. (2015) 'CLUMPAK : a program for identifying clustering modes and packaging population structure inferences across K', *Molecular Ecology*, 15, pp. 1179–1191. doi: 10.1111/1755-0998.12387.

Kotzageorgis, G. C. and Mason, C. F. (1997) 'Small mammal populations in relation to hedgerow structure in an arable landscape', *Journal of Zoology*, 242, pp. 425–434. doi: 10.1111/j.1469-7998.1997.tb03846.x.

Kumar, S. and Richardson, T. E. (2005) 'Inferring relatedness and heritability using molecular markers in radiata pine', *Molecular Breeding*, 15(1), pp. 55–64. doi:

10.1007/s11032-004-2059-4.
Latter, B. D. H. (1973) 'The sland Model of population differentiaton: a general solution', *Genetics*, 73, pp. 147–157.
Laurie, C. C., Nickerson, D. A., Anderson, A. D., Weir, B. S., Livingston, R. J., Dean, M. D., Smith, K. L., Schadt, E. E. and Nachman, M. W. (2007) 'Linkage disequilibrium in wild mice', *PLoS Genetics*, 3(8), pp. 1487–1495. doi: 10.1371/journal.pgen.0030144.
Lazzaro, B. and Little, T. (2009) 'Immunity in a variable world', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1513), pp. 15–26.
Lee YH and Kasper LH (2004) 'Immune responses of different mouse strains after challenge with equivalent lethal doses of *Toxoplasma gondii*', pp. 89–97. doi: 10.1051/parasite/200411189.
Lello, J., Boag, B., Fenton, A. and Stevenson, I. R. (2004) 'Competition and mutulaism among the gut helminths of a mamalian host', 428(April), pp. 20–24. doi: 10.1038/nature02472.1.
Lerner, A., Jeremias, P. and Matthias, T. (2016) 'The world Incidence and prevalence of autoimmune diseases is increasing', *International Journal of Celiac Disease*, 3(4), pp. 151–155. doi: 10.12691/ijcd-3-4-8.
Lesher, A., Li, B., Whitt, P., Newton, N., Devalapalli, A. P., Shieh, K., Solow, J. S. and Parker, W. (2006) 'Increased IL-4 production and attenuated proliferative and pro-inflammatory responses of splenocytes from wild-caught rats (*Rattus norvegicus*)', *Immunology and Cell Biology*, 84(4), pp. 374–382. doi: 10.1111/j.1440-1711.2006.01440.x.
Lewontin, R. C. (1962) 'Interdeme selection controlling a polymorphism in the house mouse', *The American Naturalist*, 96(887), pp. 65–78.
Lewontin, R. C. and Dunn, L. C. (1960) 'The evolutionary dynamics of a polymorphism in the house mouse', *Genetics*, 45, pp. 705–722.
Li, D., Sun, K., Zhao, Y., Lin, A., Li, S., Jiang, Y. and Feng, J. (2017) 'Polymorphism in the major histocompatibility complex (MHC class II B) genes of the Rufous-backed Bunting (*Emberiza jankowskii*)', *PeerJ*, 5, p. e2917. doi: 10.7717/peerj.2917.
Lieberman, R. and Humphrey, W. (1971) 'Association of H-2 types with genetic control of immune responsiveness to IgA allotypes in the mouse.', *Proceedings of the*

National Academy of Sciences of the United States of America, 68(10), pp. 2510–3. doi: 10.1084/jem.136.5.1222.

Lochmiller, R. L., Vestey, M. R. and Mcmurry, S. T. (1993) 'Phenotypic variation in lymphoproliferative responsiveness to mitogenic stimulation in Cotton Rats *American Society of Mammalogists*, 74(1), pp. 189–197.

Lyimo, C. M., Weigend, A., Msoffe, P. L., Eding, H., Simianer, H. and Weigend, S. (2014) 'Global diversity and genetic contributions of chicken populations from African, Asian and European Regions', *Animal Genetics*, 45(6), pp. 836–848. doi: 10.1111/age.12230.

Lynch, M. and Walsh, B. (1998) *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.

Mähler, M. and Leiter, E. H. (2002) 'Genetic and environmental context determines the course of colitis developing in IL-10-deficient mice', *Inflammatory Bowel Diseases*, 8(5), pp. 347–355. doi: 10.1097/00054725-200209000-00006.

Maizels, R. M. and Nussey, D. H. (2013) 'Into the wild: digging at immunology's evolutionary roots', *Nature Immunology*. Nature Publishing Group, 14(9), pp. 879–883. doi: 10.1038/ni.2643.

Marques, S. M., Campos, P. P., Castro, P. R., Cardoso, C. C., Ferreira, Mô. A. N. D. and Andrade, S. P. (2011) 'Genetic background determines mouse strain differences in inflammatory angiogenesis', *Microvascular Research*. Elsevier Inc., 82(3), pp. 246–252. doi: 10.1016/j.mvr.2011.08.011.

Mayer, A., Lillyt, F. and Duran-reynalst, M. L. (1980) 'Genetically dominant resistance in mice to 3-methylcholanthrene- induced lymphoma', 77(5), pp. 2960–2963.

Medina, E., Goldmann, O., Rohde, M., Lengeling, A. and Chhatwals, G. S. (2001) 'Genetic control of susceptibility to group A streptococcal infection in mice', *Journal of Infectious Diseases* , 184(7), pp. 846–852. doi: 10.1086/323292.

Medzhitov, R. and Janeway, C. J. (2000) 'Innate immune recognition: mechanisms and pathways', *Immunological Reviews*, 173(1), pp. 89–97. doi: 10.1034/j.1600-065X.2000.917309.x.

Meyer, D. and Thomson, G. (2001) 'How selection shapes variation of the human major histocompatibility complex: A review', *Annals of Human Genetics*, 1996(1), pp. 1–26. doi: 10.1046/j.1469-1809.2001.6510001.x.

Morel, L. (2004) 'Mouse models of human autoimmune diseases : essential tools that require the proper controls', *Plos Biology*, 2(8), pp. 2–5. doi: 10.1371/journal.pbio.0020241.

Mostafavi, S., Ortiz-Lopez, A., Bogue, M. A., Hattori, K., Pop, C., Koller, D., Mathis, D. and Benoist, C. (2014) 'Variation and genetic control of gene expression in primary immunocytes across inbred mouse strains.', *Journal of immunology (Baltimore, Md. : 1950)*, 193(9), pp. 4485–96. doi: 10.4049/jimmunol.1401280.

Murakami, K., Yoshikawa, S., Konishi, S., Ueno, Y., Watanabe, S. and Mizoguchi, Y. (2014) 'Evaluation of genetic introgression from domesticated pigs into the Ryukyu wild boar population on Iriomote Island in Japan', *Animal Genetics*, 45(4), pp. 517–523. doi: 10.1111/age.12157.

Nei, M. and Hughes, A. L. (1991) *Polymorphism and the evolution of the major Histocompatibility complex loci in mammals*.

Neurath, M. F., Weigmann, B., Finotto, S., Glickman, J., Nieuwenhuis, E., Iijima, H., Mizoguchi, A., Mizoguchi, E., Mudter, J., Galle, P. R., Bhan, A., Autschbach, F., Sullivan, B. M., Szabo, S. J., Glimcher, L. H. and Blumberg, R. S. (2002) 'The Transcription factor T-bet regulates mucosal T cell activation in experimental colitis and Crohn's disease', *The Journal of Experimental Medicine*, 195(9), pp. 1129–1143. doi: 10.1084/jem.20011956.

Nguyen, D. and Xu, T. (2008) 'The expanding role of mouse genetics for understanding human biology and disease', *Disease Models and Mechanisms*, 66, pp. 56–66. doi: 10.1242/dmm.000232.

Nouairia, G., Kdidi, S., Salah, R. Ben, Hammadi, H., Khorchani, T. and Yahyaoui, M. H. (2015) 'Assessing genetic diversity of three Tunisian dromedary camel (*Camelus dromedarius*) subpopulations using microsatellite markers', *Emirates Journal of Food and Agriculture*, 27(4), pp. 362–366. doi: 10.9755/ejfa.v27i4.19258.

Ohmura, K., Johnsen, A., Ortiz-Lopez, A., Desany, P., Roy, M., Besse, W., Rogus, J., Bogue, M., Puech, A., Lathrop, M., Mathis, D. and Benoist, C. (2005) 'Variation in IL-1 gene expression is a major determinant of genetic differences in arthritis aggressivity in mice', *Proceedings of the National Academy of Sciences*, 102(35), pp. 12489–12494. doi: 10.1073/pnas.0504325102.

Orrù, V. *et al.* (2013) 'Genetic variants regulating immune cell levels in health and

disease', *Cell*, 155(1). doi: 10.1016/j.cell.2013.08.041.

Packiam, M., Veit, S. J., Anderson, D. J., Ingalls, R. R. and Jerse, A. E. (2010) 'Mouse strain-dependent differences in susceptibility to *Neisseria gonorrhoeae* infection and induction of innate immune responses', *Infection and Immunity*, 78(1), pp. 433–440. doi: 10.1128/IAI.00711-09.

Paigen, K. (2003a) 'One hundred years of mouse genetics: An intellectual history. I. The classical period (1902-1980)', *Genetics*, 163(1), pp. 1–7. doi: 10.1097/00007890-197103000-00013.

Paigen, K. (2003b) 'One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981-2002).', *Genetics*, 163(4), pp. 1227–1235. doi: 10.1097/00007890-197103000-00013.

Panithanarak, T., Hauffe, H. C., Dallas, J. F., Glover, A., Ward, R. G. and Searle, J. B. (2004) 'Linkage-dependent gene flow in a house mouse chromosomal hybrid zone', *Evolution*, 58(1), pp. 184–192.

Paterson, S., Wilson, K. and Pemberton, J. M. (1998) 'Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population.', *Proceedings of the National Academy of Sciences of the United States of America*, 95(7), pp. 3714–3719. doi: 10.1073/pnas.95.7.3714.

Pedersen, A. B. and Fenton, A. (2007) 'Emphasizing the ecology in parasite community ecology', *Trends in Ecology and Evolution*, 22(3), pp. 133–139. doi: 10.1016/j.tree.2006.11.005.

Penn, D. J., Damjanovich, K. and Potts, W. K. (2002) 'MHC heterozygosity confers a selective advantage against multiple-strain infections.', *Proceedings of the National Academy of Sciences of the United States of AmericaPenn, D.J., Damjanovich, K. & Potts, W.K., 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proceedings of the National Academy of Sci*, 99(17), pp. 11260–4. doi: 10.1073/pnas.162006499.

Pepose, J. S. and Whittum-Hudson, J. A. (1987) 'An immunogenetic analysis of resistance to herpes simplex virus retinitis in inbred strains of mice', *Invest Ophthalmol Vis Sci*, 28(9), pp. 1549–1552.

Petras, M. L. (1967a) 'Studies of Natural Populations of Mus . I . Biochemical Polymorphisms and their Bearing on Breeding Structure', *Evolution*, 21(2), pp. 259-

274.

Petras, M. L. (1967b) 'Studies of natural populations of *Mus* . II . Polymorphism at the T locus', *Evolution*, 21(3), pp. 466–478.

Phifer-Rixey, M. and Nachman, M. W. (2015) 'Insights into mammalian biology from the wild house mouse *Mus musculus*', *eLife*, 4, pp. 1–13. doi: 10.7554/eLife.05959.

Piertney, S. B. and Oliver, M. K. (2006) 'The evolutionary ecology of the major histocompatibility complex', *Heredity*, 96(1), pp. 7–21. doi: 10.1038/sj.hdy.6800724.

Pitala, N., Gustafsson, L., Sendecka, J. and Brommer, J. E. (2007) 'Nestling immune response to phytohaemagglutinin is not heritable in collared flycatchers', *Biology Letters*, 3(4), pp. 418–421. doi: 10.1098/rsbl.2007.0135.

Platt, A., Vilhjálmsson, B. J. and Nordborg, M. (2010) 'Conditions under which genome-wide association studies will be positively misleading', *Genetics*, 186(3), pp. 1045–1052. doi: 10.1534/genetics.110.121665.

Pocock, M. J. ., Searle, J. B. and White, P. C. . (2004) 'Adaptations of animals to comnensal habitats : population dynamics of house mice *Mus musculus domesticus* on farms', *Journal of Animal Ecology*, 73(5), pp. 878–888.

Pocock, M. J. O., Hauffe, H. C. and Searle, J. B. (2005) 'The genus *Mus* as a model for evolutionary studies dispersal in house mice', *Biological Journal of the Linnean Society*, 84, pp. 565–583. doi: 10.1111/j.1095-8312.2005.00438.x.

Polderman, T. J. C., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M. and Posthuma, D. (2015) 'Meta-analysis of the heritability of human traits based on fifty years of twin studies', *Nature Genetics*. Nature Publishing Group, 47(7), pp. 702–709. doi: 10.1038/ng.3285.

Poltorak, A., He, X., Smirnova, I., Liu, M., Van, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Ricciardi-castagnoli, P., Layton, B. and Beutler, B. (1998) 'Defective LPS signaling in C3H / HeJ and C57BL / 10ScCr Mice : Mutations in Tlr4 gene P, *Science*, 282(5396), pp. 2085–2088.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) 'Inference of population structure using multilocus genotype data', *Genetics Society of America*, 155, pp. 945–959.

Psifidi, A., Banos, G., Matika, O., Desta, T. T., Bettridge, J., Hume, D. A., Dessie, T., Christley, R., Wigley, P., Hanotte, O. and Kaiser, P. (2016) 'Genome-wide association

studies of immune, disease and production traits in indigenous chicken ecotypes', *Genetics Selection Evolution*. BioMed Central, 48(1), pp. 1–16. doi: 10.1186/s12711-016-0252-7.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J. and Sham, P. C. (2007) 'PLINK: A tool set for whole-genome association and population-based linkage analyses', *American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795.

Queller, D. C. and Goodnight, K. F. (1989) 'Estimating relatedness using genetic markers', *Evolution*, 43(2), pp. 258–275. doi: 10.2307/2409206.

Radwan, J., Biedrzycka, A. and Babik, W. (2010) 'Does reduced MHC diversity decrease viability of vertebrate populations?', *Biological Conservation*. Elsevier Ltd, 143(3), pp. 537–544. doi: 10.1016/j.biocon.2009.07.026.

Reimer, J. D. and Petras, M. L. (1967) ' Structure of the house mouse , *Mus musculus*, in a population cage', *Journal of Mammology*, 48(1), pp. 88–99.

Reiner, G., Kliemt, D., Willems, H., Berge, T., Fischer, R., Köhler, F., Hepp, S., Hertrampf, B., Daugschies, A., Geldermann, H., Mackenstedt, U. and Zahner, H. (2007) 'Mapping of quantitative trait loci affecting resistance/susceptibility to *Sarcocystis miescheriana* in swine', *Genomics*, 89(5), pp. 638–646. doi: 10.1016/j.ygeno.2007.01.011.

Ritchie, M. D. and Van Steen, K. (2018) 'The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation', *Annals of Translational Medicine*, 6(8), pp. 157–157. doi: 10.21037/atm.2018.04.05.

Ritland, K. (1989) 'Correlated matings in the partial selfer *Mimulus guttatus*', *Evolution*, 43(4), pp. 848–859. doi: 10.2307/2409312.

Ritland, K. (1996) 'Estimators for pairwise relatedness and individual inbreeding coefficients', *Genetical Research*, 67(02), p. 175. doi: 10.1017/S0016672300033620.

Ritland, K. (2000) 'Marker-inferred relatedness as a tool for detecting heritability in nature', *Molecular Ecology*, 9(9), pp. 1195–1204. doi: 10.1046/j.1365-294X.2000.00971.x.

Ritland, K. and Ritland, C. (1996) 'Inferences about quantitative inheritance based on natural population structure in the yellow monkey flower *Mimulus Guttatus*', *Evolution*, 50(3), pp. 1074–1082.

Rivera, J. and Tessarollo, L. (2008) 'Genetic background and the dilemma of translating mouse studies to humans', *Immunity*, 28(1), pp. 1–4. doi: 10.1016/j.immuni.2007.12.008.

Rosenberg, N. A. (2004) 'DISTRUCT : A program for the graphical display of population structure', *Molecular Ecology*, 4, pp. 137–138. doi: 10.1046/j.1471-8286.2003.00566.x.

Rowe, F. P., Quy, R. J. and Swinney, T. (1987) 'Recolonization of the buildings on a farm by house mice', *Acta Theriologica*, 32(1–10), pp. 3–19.

Russell, E. (1978) *Origins of inbred mice*. New York: Academic Press.

Russi, A. E. and Brown, M. A. (2015) 'Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies', 21(7), pp. 730–738. doi: 10.1016/j.trsl.2014.08.005.

Sakaluk, S. K., Wilson, A. J., Bowers, E. K., Johnson, L. S., Masters, B. S., Johnson, B. G. P., Vogel, L. A., Forsman, A. M. and Thompson, C. F. (2015) 'Genetic and environmental variation in condition, cutaneous immunity, and haematocrit in house wrens', *BMC Evolutionary Biology*, 14(1), pp. 1–11. doi: 10.1186/s12862-014-0242-8.

Salcedo, T., Geraldes, A. and Nachman, M. W. (2007) 'Nucleotide variation in wild and inbred mice', *Genetics*, 177(4), pp. 2277–2291. doi: 10.1534/genetics.107.079988.

Searle, J. B., Jones, C. S., Gündüz, I., Scascitelli, M., Jones, E. P., Herman, J. S., Rambau, R. V., Noble, L. R., Berry, R. J., Giménez, M. D. and Jóhannesdóttir, F. (2009) 'Of mice and (Viking?) men: phylogeography of British and Irish house mice.', *The Royal Society*, 276(1655), pp. 201–7. doi: 10.1098/rspb.2008.0958.

Selander, R. K. (1970) 'Behavior and genetic variation in natural populations', *American Zoology*, 10(1), pp. 53–66.

Sellers, L. A., Long, R. F., Jay-Russell, M. T., Li, X., Atwill, E. R., Engeman, R. M. and Baldwin, R. A. (2018) 'Impact of field-edge habitat on mammalian wildlife abundance, distribution, and vectored foodborne pathogens in adjacent crops', *Crop Protection*, pp. 1–11. doi: 10.1016/j.cropro.2018.02.005.

Sellers, R. S., Clifford, C. B., Treuting, P. M. and Brayton, C. (2012) 'Immunological variation between inbred laboratory mouse strains: points to consider in phenotyping genetically immunomodified mice', *Veterinary Pathology*, 49(1), pp. 32–43. doi: 10.1177/0300985811429314.

Singleton, G. R. and Krebs, C. J. (2006) 'The secret world of mice; a geographic distribution', in *The House Mouse in Biomedical Research: History, Wild Mice, and Genetics*, pp. 26–27.

Sitepu, P. and Dobson, C. (1982) 'Genetic control of resistance to infection with Nematospiroides dubius in mice: selection of high- and low-immune response populations of mice.', *Parasitology*. University of Bristol Library, 85(1982), pp. 73–84.

Smith, G. S., Walford, R. L. and Mickey, M. R. (1973) 'Lifespan and Incidence of cancer and other diseases in selected long-Lived inbred mice and their F1 hybrids', *JNCI: Journal of the National Cancer Institute*, 50(5), pp. 1195–1213. doi: 10.1093/jnci/50.5.1195.

Smith, J. A., Wilson, K., Pilkington, J. G. and Pemberton, J. M. (1999) 'Heritable variation in resistance to gastro-intestinal nematodes in an unmanaged mammal population', *Proceedings of the Royal Society B: Biological Sciences*, 266(1425), pp. 1283–1290. doi: 10.1098/rspb.1999.0776.

Sommer, S. (2005) 'The importance of immune gene variability (MHC) in evolutionary ecology and conservation', *Frontiers in Zoology*, 2, pp. 1–18. doi: 10.1186/1742-9994-2-16.

Stephan, K., Smirnova, I., Jacque, B. and Poltorak, A. (2007) 'Genetic analysis of the innate immune responses in wild-derived inbred strains of mice', *European journal of immunology*, 37(1), pp. 212–223. doi: 10.1002/eji.200636156.

Sugg, D. W., Kennedy, M. L. and Heidt, G. A. (1990) 'Genetic variation in the Texas mouse , *Peromyscus attwateri*', *Journal of Mammology*, 71(3), pp. 309–317.

Suzuki, H., Nunome, M., Kinoshita, G., Aplin, K. P., Vogel, P., Kryukov, A. P., Jin, M. L., Han, S. H., Maryanto, I., Tsuchiya, K., Ikeda, H., Shiroishi, T., Yonekawa, H. and Moriwaki, K. (2013) 'Evolutionary and dispersal history of Eurasian house mice *Mus musculus* clarified by more extensive geographic sampling of mitochondrial DNA', *Heredity*. Nature Publishing Group, 111(5), pp. 375–390. doi: 10.1038/hdy.2013.60.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) 'MEGA6 :

Molecular Evolutionary Genetics Analysis Version 6 . 0', *Molecular Biology and Evolution*, 30(12), pp. 2725–2729. doi: 10.1093/molbev/mst197.
R Core Team. (2013) 'R: A language and environment for statistical computing. R Foundation for Statistical Computing'. Vienna.
RStudio Team. (2015) 'RStudio: integrated development for R. RStudio, Inc.',. Boston, MA.
Telfer, S., Lambin, X., Birtles, R., Beldomenico, P., Paterson, S. and Begon, M. (2010) 'Species interactions in a parasite community drive infection risk in a wildlife population', 330(6001), pp. 243–246.
Thia, K. Y. T. and Trapani, J. A. (2007) 'The granzyme B gene is highly polymorphic in wild mice but essentially invariant in common inbred laboratory strains', *Tissue Antigens*, 70(3), pp. 198–204. doi: 10.1111/j.1399-0039.2007.00872.x.
Thomas, S. C., Coltman, D. W. and Pemberton, J. M. (2002) 'The use of marker-based relationship information to estimate the heritability of body weight in a natural population: A cautionary tale', *Journal of Evolutionary Biology*, 15(1), pp. 92–99. doi: 10.1046/j.1420-9101.2002.00372.x.
Thompson-Crispi, K. A., Sargolzaei, M., Ventura, R., Abo-Ismail, M., Miglior, F., Schenkel, F. and Mallard, B. A. (2014) 'A genome-wide association study of immune response traits in Canadian Holstein cattle', *BMC Genomics*, 15(1). doi: 10.1186/1471-2164-15-559.
Trowsdale, J. and Parham, P. (2004) 'Defense strategies and immunity-related genes', *European Journal of Immunology*, 34(1), pp. 7–17. doi: 10.1016/j.euromechsol.2009.10.002.
Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H., Mittl, P. and Ratberg, L. (2013) 'Polymorphisms at the innate immune receptor TLR2 are associated with borrelia Infection in a wild rodent population', *Proceedings. biological sciences / The Royal society*, 280(1759), p. 20130364.
Turner, A (2010) 'Immunogenetics and Polymorphism in a Natural Population of Field voles (Microtus agrestis)', *PLoS Genetics*.
Turner, A., Begon, M., Jackson, J., Bradley, J. and Paterson, S. (2011) 'Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population', *PLoS genetics*, 7(10), p. e1002343. doi:

10.1371/journal.pgen.1002343.
Uzzaman, M. R., Edea, Z., Bhuiyan, M. S. A., Walker, J., Bhuiyan, A. K. F. H. and Kim, K. S. (2014) 'Genome-wide single nucleotide polymorphism analyses reveal genetic diversity and structure of wild and domestic cattle in Bangladesh', *Asian-Australasian Journal of Animal Sciences*, 27(10), pp. 1381–1386. doi: 10.5713/ajas.2014.14160.
Valdar, W., Solberg, L. C., Gauguier, D., Cookson, W. O., Rawlins, J. N. P., Mott, R. and Flint, J. (2006) 'Genetic and environmental effects on complex traits in mice', *Genetics*, 174(2), pp. 959–984. doi: 10.1534/genetics.106.060004.
Vilhjálmsson, B. J. and Nordborg, M. (2013) 'The nature of confounding in genome-wide association studies', *Nature Reviews Genetics*, 14(1), pp. 1–2. doi: 10.1038/nrg3382.
Viney, M and Riley, E. (2014) 'From Immunology to eco-immunology: more than a new name'. *Eco-immunolo*gy, pp1-19, doi: 10.1007/978-94-017-8712-3_1
Viney, M., Lazarou, L. and Abolins, S. (2015) 'The laboratory mouse and wild immunology', *Parasite Immunology*, 37(5), pp. 267–273. doi: 10.1111/pim.12150.
Viney, M., Riley, E. and Buchanan, K. (2005) 'Optimal immune responses: immunocompetence revisited', *Trends in Ecology & Evolution*, 20(12), pp. 665–669. doi: 10.1016/j.tree.2005.10.003.
Visscher, P. M., Hill, W. G. and Wray, N. R. (2008) 'Heritability in the genomics era - concepts and misconceptions', *Nature Reviews Genetics*, 9(4), pp. 255–266. doi: 10.1038/nrg2322.
Walford, R. L., Liu, R. K., Gerbase-Delima, M., Mathies, M. and Smith, G. S. (1973) 'Longterm dietary restriction and immune function in mice: response to sheep red blood cells and to mitogenic agents', *Mechanisms of Ageing and Development*, 2, pp. 447–454. doi: 10.1016/0047-6374(73)90035-3.
Waterson, R. H. and Consortium, M. G. S. (2002) 'Initial sequencing and comparative analysis of the mouse genome', *NATURE*, 420, pp. 520–562.
Wells, J. C. K. and Stock, J. T. (2011) 'Re-examining heritability: genetics, life history and plasticity', *Trends in Endocrinology and Metabolism*. Elsevier Ltd, 22(10), pp. 421–428. doi: 10.1016/j.tem.2011.05.006.
Whalund, S. (1928) 'Zusammensetzung von populationen und korrelationerscheinungen vom standpunkt der vererbungslehre aus betrachtet',

Hereditas, 11, pp. 65–106.

Williams-Blangero, S., Correa-Oliveira, R., Vandeberg, J. L., Subedi, J., Upadhayay, R. P., Rai, D. R., Jha, B. and Blangero, J. (2004) 'Genetic influences on plasma cytokine variation in a parasitized population', *Human Biology*, 76(4), pp. 515–525. doi: 10.1353/hub.2004.0062.

Williams, N. M. and Timoney, P. J. (1994) 'Variation in susceptibility of ten mouse strains to infection with a strain of *Ehrlichia risticii*', *Journal of Comparative Pathology*, 110(2), pp. 137–143. doi: 10.1016/S0021-9975(08)80185-5.

Wilson, A. J., McDonald, G., Moghadam, H. K., Herbinger, C. M. and Fergeson, M. M. (2003) 'Marker-assisted estimation of quantitative genetic parameters in rainbow trout, *Oncorhynchus mykiss*', *Genetical Research*, 81(2), doi: 10.1017/S0016672302006055.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B. and Nussey, D. H. (2010) 'An ecologist's guide to the animal model', *Journal of Animal Ecology*, 79(1), pp. 13–26. doi: 10.1111/j.1365-2656.2009.01639.x.

Wilson, G. M., Bussche, R. A. Van Den, Kennedy, P. K., Gunn, A., Poole, K., Gunn, A. and Poole, K. I. M. (2000) 'Genetic variability of wolverines (*Gulo gulo*) from the Northwest Territories , Canada : conservation implications', *Journal of Mammology*, 81(1), pp. 186–196.

Wright, S. (1922) 'Coefficients of inbreeding and reationship', *Amreican Naturalist*, 56, pp. 330–338.

Wright, S. (1946) 'Isolation by distance under diverse systems of mating', *Genetics*, 31(1), pp. 39–59.

Xavier, R. J. and Rioux, J. D. (2008) 'Genome-wide association studies: a new window into immune-mediated diseases', *Nature Reviews Immunology*, 8(8), pp. 631–643. doi: 10.1038/nri2361.

Yang, J., Benyamin, B., McEvoy, B., Gordon, S., Henders, A., Nyholt, D., Maden, P., Heath, A., Martin, N., Montgomery, G., Goddard, M. and Visscher, M. (2010) 'Common SNPs explain a large proportion of heritability for human height', *Nature genetics*, 42(7), pp. 565–569. doi: 10.1038/ng.608.Common.

Yang, J., Lee, S. H., Goddard, M. E. and Visscher, P. M. (2011) 'GCTA: A tool for genome-wide complex trait analysis', *American Journal of Human Genetics*. The

American Society of Human Genetics, 88(1), pp. 76–82. doi: 10.1016/j.ajhg.2010.11.011.

Yao, Y.-F., Dai, Q.-X., Li, J., Ni, Q.-Y., Zhang, M.-W. and Xu, H.-L. (2014) 'Genetic diversity and differentiation of the Rhesus macaque (*Macaca mulatta*) Population in Western Sichuan, China, based on the second exon of the major histocompatibility complex class II DQB (MhcMamu-DQB1) Alleles.', *BMC evolutionary biology*, 14(1), p. 130. doi: 10.1186/1471-2148-14-130.